Protocol for the Systematic Review of the Health Effects of Phthalate Exposure

November 2018

TABLE OF CONTENTS

[TOC \o "1-3" \h \z \u]

LIST OF TABLES

[TOC \h \z \c "Table"]

LIST OF APPENDIX TABLES

[TOC \h \z \c "Appendix Table"]

1. BACKGROUND, SCOPING, AND PROBLEM FORMULATION SUMMARY

1.1. Exposure

Phthalates are a group of alkyl diesters of phthalic acid that are used in a wide variety of consumer products including cosmetics, personal-care products, pharmaceuticals, medical devices, children's toys, food packaging, and cleaning and building materials. Because of the widespread use, humans are exposed to mixtures of multiple phthalates across all life stages [ADDIN EN.CITE ADDIN EN.CITE.DATA].

The routes by which humans are exposed to phthalates and the magnitude of individual phthalate exposures have changed over time as the quantities and uses of the various phthalates have changed. Environmental concentrations of phthalates are typically the highest in house dust, and they may be present in food due to the use of phthalates in packaging and food preparation materials. For most phthalates, food ingestion is the dominant pathway of exposure, with dust exposures (ingestion and dermal contact) and inhalation also being important in some circumstances. Infant and toddler exposures may also occur through ingestion due to mouthing and playing with plastic toys that contain phthalates [ADDIN EN.CITE <EndNote><Cite><Author>Wormuth</Author><Year>2006</Year><RecNum>2</RecNum><DisplayText>(Wormuth et al., 2006)</DisplayText><record><rec-number>2</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509460009">2</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Wormuth, M.</author><author>Scheringer, M.</author><author>Vollenweider, M.</author><author>Hungerbuhler, K.</author></authors></contributors><titles><title>What are the sources of exposure to eight frequently used phthalic acid esters in Europeans?</title><secondary-title>Risk Analysis</secondary-title><alt-title>Risk Anal</alt-title></titles><periodical><full-title>Risk Analysis</full-title><abbr-1>Risk Anal</abbr-1></periodical><alt-periodical><full-title>Risk Analysis</full-title><abbr-1>Risk Anal</abbr-1></alt-periodical><pages>803-824</pages><volume>26</volume><number>3</number><dates><year>2006</year></dates><isbn>ISSN 0272-4332
EISSN 1539-6924</isbn><accession-num>16834635</accession-num><label>680214</label><urls><related-urls><url>http://dx.doi.org/10.1111/j.1539-6924.2006.00770.</url></related-urls></urls><electronic-resource-num>10.1111/j.1539-6924.2006.00770.</electronic-resource-num><language>English</language></record></Cite></EndNote>].

1.2. Concerns for phthalate toxicity

Concerns over human exposure to phthalates have largely centered on male reproductive toxicity. In male rats, it has been established that gestational exposure to certain phthalates produces a phenotype known as "phthalate syndrome", which is characterized by cryptorchidism,

reduced anogenital distance, female-like nipple retention, hypospadias, and malformations of the epididymis, vas deferens, seminal vesicles, and prostate [ADDIN EN.CITE ADDIN EN.CITE.DATA]. Phthalates produce this phenotype through multiple modes of action: inhibition of testosterone production; inhibition of insulin-like-3 hormone, which regulates transabdominal testicular descent; and disruption of Sertoli cells and germ cell development, which occurs via an androgen-independent mode of action (MOA) [ADDIN EN.CITE ADDIN EN.CITE.DATA].

Based on concerns over effects on the developing male reproductive tract, the Consumer Product Safety Commission (CPSC) has acted to permanently ban certain antiandrogenic phthalates in any amount greater than 0.1% in children's toys. Three phthalates were permanently banned under the Consumer Product Safety Improvement Act of 2008 [diethylhexyl phthalate (DEHP), dibutyl phthalate (DBP), and benzylbutyl phthalate (BBP)], and a final rule was issued in 2017 to expand the permanent ban to include five additional phthalates [diisononyl phthalate (DINP), diisobutyl phthalate (DIBP), di-n-pentyl phthalate, di-n-hexyl phthalate, and dicyclohexyl phthalate] [16 CFR Part 1307 (2017)]. The National Academy of Sciences (NAS) also conducted a recent systematic review to characterize the low dose effects of phthalate exposure on male reproductive development,[1] focusing on three outcomes that have known association with phthalate exposure: decreased testosterone, anogenital distance, and hypospadias [ADDIN EN.CITE <EndNote><Cite><Author>NAS</Author><Year>2017</Year><RecNum>47</RecNum><DisplayText>(NAS, 2017)</DisplayText><record><rec-number>47</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509470742">47</key></foreign-keys><ref-type name="Book">6</ref-type><contributors><authors><author>NAS,</author></authors></contributors><titles><title>Application of systematic review methods in an overall strategy for evaluating low-dose toxicity from endocrine active chemicals</title><secondary-title>Consensus Study Report</secondary-

[1] The NAS systematic review of phthalates and male reproductive tract development included the following phthalate diesters and their corresponding monoester or oxidative metabolites: BBP, DBP, DEP, DEHP, DIBP, DINP, diisooctyl phthalate, dimethyl phthalate, di-n-octyl phthalate, diisodecyl phthalate, and dipentyl phthalate [ADDIN EN.CITE <EndNote><Cite><Author>NAS</Author><Year>2017</Year><RecNum>47</RecNum><DisplayText>(NAS, 2017)</DisplayText><record><rec-number>47</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509470742">47</key></foreign-keys><ref-type name="Book">6</ref-type><contributors><authors><author>NAS,</author></authors></contributors><titles><title>Application of systematic review methods in an overall strategy for evaluating low-dose toxicity from endocrine active chemicals</title><secondary-title>Consensus Study Report</secondary-title></titles><dates><year>2017</year></dates><pub-location>Washington, D.C.</pub-location><publisher>The National Academies Press</publisher><label>3982546</label><urls><related-urls><url>http://dx.doi.org/10.17226/24758</url></related-urls></urls><electronic-resource-num>10.17226/24758</electronic-resource-num><language>English</language></record></Cite></EndNote>].

title></titles><dates><year>2017</year></dates><pub-location>Washington, D.C.</pub-location><publisher>The National Academies Press</publisher><label>3982546</label><urls><related-urls><url>http://dx.doi.org/10.17226/24758</url></related-urls></urls><electronic-resource-num>10.17226/24758</electronic-resource-num><language>English</language></record></Cite></EndNote>].

In addition to male reproductive toxicity, experimental animal studies also indicate that phthalate exposure may produce effects such as decreased maternal progesterone levels, leading to spontaneous abortions in mid-gestation [ADDIN EN.CITE <EndNote><Cite><Author>Gray</Author><Year>2006</Year><RecNum>7</RecNum><DisplayText>(Gray et al., 2006)</DisplayText><record><rec-number>7</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509460657">7</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Gray, L. E., Jr.</author><author>Laskey, J.</author><author>Ostby, J.</author></authors></contributors><titles><title>Chronic di-n-butyl phthalate exposure in rats reduces fertility and alters ovarian function during pregnancy in female Long Evans hooded rats</title><secondary-title>Toxicological Sciences</secondary-title><alt-title>Toxicol Sci</alt-title></titles><periodical><full-title>Toxicological Sciences</full-title><abbr-1>Toxicol Sci</abbr-1></periodical><alt-periodical><full-title>Toxicological Sciences</full-title><abbr-1>Toxicol Sci</abbr-1></alt-periodical><pages>189-195</pages><volume>93</volume><number>1</number><dates><year>2006</year></dates><isbn>ISSN 1096-6080
EISSN 1096-0929</isbn><accession-num>16763070</accession-num><label>673276</label><urls><related-urls><url>http://dx.doi.org/10.1093/toxsci/kfl035</url></related-urls></urls><electronic-resource-num>10.1093/toxsci/kfl035</electronic-resource-num><language>English</language></record></Cite></EndNote>]; and structural abnormalities in developing fetuses [ADDIN EN.CITE ADDIN EN.CITE.DATA]. Epidemiological studies provide evidence of an association between phthalate exposure and other health outcomes including Type 2 diabetes [ADDIN EN.CITE <EndNote><Cite><Author>Kuo</Author><Year>2013</Year><RecNum>11</RecNum><DisplayText>(Kuo et al., 2013)</DisplayText><record><rec-number>11</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509460815">11</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Kuo, C. C.</author><author>Moon, K.</author><author>Thayer, K. A.</author><author>Navas-Acien, A.</author></authors></contributors><titles><title>Environmental chemicals and type 2 diabetes: an updated systematic review of the epidemiologic evidence</title><secondary-title>Current Diabetes Reports</secondary-title><alt-title>Curr Diab Rep</alt-

title></titles><periodical><full-title>Current Diabetes Reports</full-title><abbr-1>Curr Diab Rep</abbr-1></periodical><alt-periodical><full-title>Current Diabetes Reports</full-title><abbr-1>Curr Diab Rep</abbr-1></alt-periodical><pages>831-849</pages><volume>13</volume><number>6</number><dates><year>2013</year></dates><isbn>ISSN 1534-4827
EISSN 1539-0829</isbn><accession-num>24114039</accession-num><label>2088454</label><work-type>Review</work-type><urls><related-urls><url>http://dx.doi.org/10.1007/s11892-013-0432-6</url></related-urls></urls><electronic-resource-num>10.1007/s11892-013-0432-6</electronic-resource-num><language>English</language></record></Cite></EndNote>] and neurodevelopmental effects [ADDIN EN.CITE <EndNote><Cite><Author>Ejaredar</Author><Year>2015</Year><RecNum>12</RecNum><DisplayText>(Ejaredar et al., 2015)</DisplayText><record><rec-number>12</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509460868">12</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Ejaredar, M.</author><author>Nyanza, E. C.</author><author>Ten Eycke, K.</author><author>Dewey, D.</author></authors></contributors><titles><title>Phthalate exposure and childrens neurodevelopment: A systematic review</title><secondary-title>Environmental Research</secondary-title><alt-title>Environ Res</alt-title></titles><periodical><full-title>Environmental Research</full-title><abbr-1>Environ Res</abbr-1></periodical><alt-periodical><full-title>Environmental Research</full-title><abbr-1>Environ Res</abbr-1></alt-periodical><pages>51-60</pages><volume>142</volume><dates><year>2015</year></dates><isbn>ISSN 0013-9351
EISSN 1096-0953</isbn><accession-num>26101203</accession-num><label>3045520</label><work-type>Review</work-type><urls><related-urls><url>http://dx.doi.org/10.1016/j.envres.2015.06.014</url></related-urls></urls><electronic-resource-num>10.1016/j.envres.2015.06.014</electronic-resource-num><language>English</language></record></Cite></EndNote>]. These effects have been less widely studied compared to the male reproductive toxicity of phthalates, and therefore may not be as well characterized.

1.3. Problem formulation

A collection of systematic reviews was designed to identify the range of health effects associated with exposure to the phthalates listed in Table 1, including emerging health outcomes that were not covered by the recent Consumer Product Safety Commission (CPSC) and National Academies of Science (NAS) reviews.

- For epidemiology studies, the systematic review was designed to target any heath outcome associated with exposure to the phthalates listed in Table 1 (DIBP, DEP, DINP, BBP, DBP, DEHP).

- For animal studies, the systematic review focused on six health outcome categories: male reproductive, female reproductive, developmental, liver, kidney, and cancer. These categories were selected because a preliminary assessment of the extent of research in these areas indicated the availability of data to support hazard synthesis. Reviews were performed for two phthalates: DIBP, which has generally been found to be antiandrogenic, and DEP, which has generally been found not to share the antiandrogenic MOA. Additional details of these reviews are provided in this protocol.

Table [SEQ Table * ARABIC]. Nomenclature and abbreviations of phthalates and phthalate metabolites

Parent compound, abbreviation(s)[a]		**Major metabolite(s) in humans, abbreviation(s)**[b]	
Diethyl phthalate	**DEP**	Monoethyl phthalate	**MEP**
Dibutyl phthalate, *Di-n-butyl phthalate,* *Butyl phthalate,* *n-Butyl phthalate,* *Bis(n-butyl) phthalate*	**DBP,** DnBP, DnBuP	Monobutyl phthalate *Mono-n-butyl phthalate*	**MBP** MnBP, MnBuP
Diisobutyl phthalate	DiBP, **DIBP**	Monoisobutyl phthalate	**MiBP**, MiBuP
Butyl benzyl phthalate, *Benzyl butyl phthalate,* *n-Butyl benzyl phthalate*	BBzP, **BBP**	Monobenzyl phthalate	**MBzP**, MBeP
Di(2-ethylhexyl) phthalate, *Dioctyl phthalate,* *Bis(2-ethylhexyl) phthalate*	**DEHP,** DOP	Mono(2-ethylhexyl) phthalate[b]	**MEHP**
		Mono(2-ethyl-5-hydroxyhexyl) phthalate	**MEHHP**, 5-OH-MEHP
		Mono(2-ethyl-5-oxohexyl) phthalate	**MEOHP**, 5oxo-MEHP,
		Mono(2-ethyl-5-carboxypentyl) phthalate	**MECPP**, 5cx-MEPP
Diisononyl phthalate	DiNP, **DINP**	Monoisononyl phthalate[b]	MiNP, **MINP,** MNP
		Mono(hydroxylisononyl) phthalate	**MHiNP**, OH-MiNP, 7OH-MMeOP
		Mono(oxoisononyl) phthalate	**MOiNP**, oxo-MiNP, 7oxo-MMeOP
		Mono(carboxyisooctyl) phthalate	**MCiOP**, MCOP, cx-MiNP, 7cx-MMeHP

[a]Bold indicates abbreviation used in this report.
[b]Primary monoester for long-chain phthalate.

2. OVERALL OBJECTIVES, SPECIFIC AIMS AND PECO

The overall objective of these systematic reviews is to evaluate the hazards of individual phthalates by conducting a systematic review of existing epidemiological and toxicological literature, including consideration of relevant mechanistic evidence.

2.1. Specific aims

- Identify epidemiological and experimental animal literature reporting the effects of exposure to phthalates on the health outcomes listed in the PECO (Population, Exposure, Comparator, Outcome) (Table 2).

- Identify studies reporting *in vitro* and other types of mechanistic evidence. An iterative approach will be used to determine which *in vitro* and other types of mechanistic studies are most important to also summarize, based on factors such as directness or relevance of the model systems, concentrations tested, and robustness of the evidence in humans and animals.

- Conduct study evaluation for individual epidemiological and animal health effect studies. Evaluation considered domains relating to reporting quality, risk of bias, and sensitivity. For discussion of these concepts, see [HYPERLINK \l "_ENREF_37" \o "Rooney, 2016 #288"] and [HYPERLINK \l "_ENREF_10" \o "Cooper, 2016 #32"]

- Extract data on relevant health outcomes from included epidemiological and experimental animal studies. Data are not extracted from studies considered uninformative following study evaluation and thus are not considered further in the analysis.

- Synthesize the evidence across studies assessing similar health outcomes using a narrative approach or meta-analysis (if appropriate) and evaluate sources of heterogeneity.

- For each phthalate and health outcome, express confidence in conclusions from across studies (or subsets of studies) within human and animal evidence streams according to one of five statements: 1. Robust, 2. Moderate, 3. Slight 4. Indeterminate or 5. Compelling evidence of no effect. Each evidence stream will be evaluated separately.

- Characterize uncertainties and identify key data gaps and research needs, e.g., related to limitations of the evidence base, limitations of the systematic review, consideration of dose-relevance and pharmacokinetic differences when extrapolating findings from animal studies to human exposure levels.

2.2. Assessment approach

Epidemiological and animal data will be reviewed and synthesized separately for each individual phthalate.

Table [SEQ Table * ARABIC]. PECO (Populations, Exposures, Comparators, Outcomes)

PECO	Evidence
Population	**Human:** Any population (children, general population, occupational, high exposure from an environmental source). The following study designs were considered potentially informative: controlled exposure, cohort, case-control, or cross-sectional.
	Animal: Non-human mammalian animal species (whole organism) of any lifestage (including preconception, in utero, lactation, peripubertal and adult stages). An iterative approach was used to prioritize evidence from non-mammalian model systems (e.g., fish, amphibians, birds, *C. elegans,* etc.) based on likelihood to impact evidence synthesis conclusions. Evidence from non-mammalian model systems was preliminarily tagged during title/abstract screening as "Studies with Supporting Data".
	Mechanistic: Human or animal cells, tissues, or biochemical reactions (e.g., ligand binding assays) with *in vitro* exposure regimens; bioinformatics pathways of disease analysis; or high throughput screening data. An iterative approach was used to prioritize mechanistic studies for analysis based on likelihood to impact evidence synthesis conclusions. During title/abstract screening, mechanistic studies were preliminarily tagged as "Studies with Supporting Data".
Exposure	**Human:** Exposure to one or more of the phthalates listed in Table 1, as singular compounds or as mixtures, as determined by: • Measured concentration in contact medium (e.g., air, dust) • Biomarkers of exposure (e.g., urinary metabolite levels of phthalates) Occupation involving exposure to phthalates (e.g., plastics manufacture); knowledge of specific contamination sites or accidental exposure.
	Animal: Exposure to any administered dose or concentration of DIBP or DEP or their major metabolites (MIBP or MEP; see Table 1) as singular compounds. Exposure routes may include any oral, inhalation, or dermal exposures.
Comparator	**Human:** A comparison population exposed to lower levels (or no exposure/exposure below detection levels).
	Animal (and mechanistic): Exposed to vehicle-only treatment or untreated control.
Outcomes	**Human:** Any examination of human health effects.
	Animal: Any examination of the following effects: male reproductive, female reproductive, developmental, liver, kidney, cancer.

3. LITERATURE SEARCH AND SCREENING STRATEGIES

3.1. Literature search strategies

The literature search strategy consisted of a broad search of online scientific databases, casting a wide net in order to identify all potentially pertinent studies. Animal studies were identified by conducting separate literature searches for each phthalate (DIBP or DEP), whereas epidemiology studies were identified by conducting a single broad literature search on all six phthalates (DIBP, DEP, DINP, BBP, DBP, DEHP). This strategy for the epidemiological literature search provides advantages over separate searches for epidemiology literature on individual phthalates because epidemiology studies frequently examine multiple phthalate exposures in a single study (e.g. metabolites of several different phthalates). A single search of all phthalates reduces the likelihood of screening the same study multiple times. Additionally, indexing terms and abstracts may not include a comprehensive list of all phthalates examined in a study, so literature search for individual phthalates have the potential to miss studies and introduce bias in the selection process. This is particularly true because "negative" studies (i.e. studies that did not demonstrate an association between exposure and outcome) are more likely than "positive" studies to be missed in searches based on single phthalates.

The search strategies to identify all literature were developed in consultation with an information specialist and are presented in Appendix 1. The following databases were searched:

- PubMed (epidemiology and animal studies)
- Web of Science (epidemiology and animal studies)
- Toxline (epidemiology and animal studies)
- Toxic Substances Control Act Test Submissions (TSCATS2) (animal studies)
- Toxcenter (animal studies for DEP)

Non-date-limited searches with no language restrictions were applied. The initial search was conducted as early as March 2012 and was followed by literature search updates every 6–12 months through July 2017. Literature searching were conducted using the EPA's Health and Environmental Research Online (HERO) database.[2]

Additional relevant literature not found through database searching was identified through:

---

[2] HERO ([HYPERLINK "https://hero.epa.gov/hero/"]) is a database of scientific studies and other references used to develop EPA's risk assessments aimed at understanding the health and environmental effects of pollutants and chemicals. It is developed and managed in EPA's Office of Research and Development (ORD) by the National Center for Environmental Assessment (NCEA). The database includes more than 1,400,000 scientific articles from the peer-reviewed literature. New studies are added continuously to HERO.

- Searching citations from key references (including review articles), including "backward" (to identify articles cited by key studies) and "forward" (to identify articles that cite the key study) searches.

- Manual search of citations from key regulatory documents, which were identified through a search of online chemical assessment-related websites.

- Search of references from previous assessments by the EPA's Integrated Risk Information System (IRIS) and/or references that had been previously added to the HERO project page for these chemicals.

- Writing to 94 corresponding authors of primary research studies identified through the June 2014 literature update (done in November 2014). These studies covered eight topic areas in epidemiology: diabetes, female reproductive effects, male reproductive effects, sexual differential effects, neurodevelopment, obesity, thyroid effects, and immune (allergy and asthma) effects. The correspondence (sent by email) included a list of the identified studies in the topic area, and asked if the author knew of any studies, including unpublished studies, that had not been identified.

3.2. Unpublished data

Unpublished data from personal author communication can supplement a peer-reviewed study, as long as the information is made publicly available.

3.3. Screening Process

Studies that complied with the criteria specified in the PECO (Table 2) were eligible for inclusion, while those that did not meet these criteria were excluded. In addition to these criteria, the following exclusion criteria were applied:

- Records that do not contain original data, such as reviews, editorials, or commentaries.

- Studies that have not been peer-reviewed (e.g., conference abstracts, technical reports, working papers from research groups or committees, and white papers).

Studies were screened for inclusion using a structured form in Microsoft Excel spreadsheets. Following a pilot phase to calibrate screening guidance, two screeners independently conducted a title and abstract screen of the search results to identify records that appear to meet the PECO eligibility criteria. Records not excluded based on the title and abstract were moved forward for full-text review. For citations with no abstract, articles were screened based on all or some of the following: title relevance (title should indicate clear relevance), page numbers (articles two pages in length or less are assumed conference reports, editorials, or letters), and PubMed MeSH (Medical Subject Headings). Screening conflicts were resolved by discussion among the primary screeners with consultation by a third reviewer or technical advisor (if needed) to resolve any remaining disagreements. Assessment of eligibility status of non-English studies was facilitated

by native-language speakers at SRC, Inc. Supporting information that is not directly applicable to the PECO (e.g., ADME, exposure characteristics) was tracked during the screening process. Conflict resolution is not required during the screening process to identify supporting information, i.e., tagging by a single screener is sufficient to identify the study as potential supportive information.

Full-text copies of potentially relevant records identified from title and abstract screening were retrieved, stored in the HERO database, and independently assessed by two screeners to confirm eligibility according to the PECO criteria. Non-English studies were prioritized for translation if they met PECO criteria and were considered likely to contribute unique information to the evidence synthesis. Screening conflicts were resolved by discussion among the primary screeners with consultation by a third reviewer or technical advisor as needed to resolve any remaining disagreements.

The included and excluded studies for the epidemiological assessment and for the individual phthalates assessments are posted on the respective project pages for these assessments in the HERO database:

- **Epidemiological assessment (DIBP, DEP, DINP, BBP, DBP, DEHP):** [HYPERLINK "https://hero.epa.gov/hero/index.cfm/project/page/project_id/2245"]
- **DIBP**: [HYPERLINK "https://hero.epa.gov/hero/index.cfm/project/page/project_id/2320"]
- **DEP**: [HYPERLINK "https://hero.epa.gov/hero/index.cfm/project/page/project_id/1097"]

3.4. Multiple publications of the same data

Multiple publications with overlapping data for the same study (e.g., publications reporting subgroups, additional outcomes or exposures outside the scope of an evaluation, or longer follow-up) were identified by examining author affiliations, study designs, cohort name, enrollment criteria, and enrollment dates. If necessary, study authors were contacted to clarify any uncertainty about the independence of two or more articles. In instances where multiple publications were available for the same study, one study was selected to use as the primary, and all others were considered as secondary publications with annotation as being related to the primary record during data extraction. The primary study was generally the publication with the longest follow-up, or for studies with equivalent follow-up periods, the study with the largest number of cases or the most recent publication date. Relevant data was included from all publications of a study, although if the same outcome is reported in more than one publication, the duplicate data was excluded.

3.5. Literature surveys and summary-level inventories

During title/abstract or full-text level screening, studies were categorized (or "tagged") based on features such as evidence stream (human, animal, in vitro, in silico), health outcomes and/or endpoint measures included in the study, or type of mechanistic information (mechanistic, PBPK, ADME, etc.).

For epidemiology studies, a literature inventory was created to develop summary-level, sortable lists that include some basic study design information (e.g., study population, biomarkers analyzed, media for exposure measure, parent phthalates and phthalate metabolites measured, health outcomes, etc.) and administrative data (e.g., full citation, HERO ID, link to HERO, search date, etc.) using Microsoft Access. This literature inventory facilitated subsequent review of individual studies or sets of studies by topic-specific experts.

3.6. Tracking study eligibility and reporting the flow of information

The main reason for exclusion at the full-text-review stage was annotated and reported in the study flow diagram within the assessment. Commonly used categories for exclusion include the following: (1) not relevant to PECO; (2) is a review, commentary, or letter with no original data; (3) is a conference abstract (and the criteria for including unpublished data, described above, are not met); or (4) unable to obtain full-text. Reasons for exclusions identified during data extraction or study evaluation, e.g., key deficiencies in reporting quality or concerns for bias/insensitivity, can be annotated at the full-text review level.

4. STUDY EVALUATION (REPORTING, RISK OF BIAS, AND SENSITIVITY) STRATEGY

For each study selected for inclusion, the quality and informativeness of the evidence was rated by evaluating for metrics related to reporting quality, risk of bias, and sensitivity. Reporting quality refers to how well the study authors communicated the details of the methods and results. Risk of bias, sometimes referred to as internal validity, is the extent to which the design or conduct of a study may alter the ability to provide accurate (unbiased) evidence to support the relationship between exposure and effects [ADDIN EN.CITE
<EndNote><Cite><Author>Higgins</Author><Year>2011</Year><RecNum>286</RecNum><DisplayText>(Higgins, 2011)</DisplayText><record><rec-number>286</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1522683248">286</key></foreign-keys><ref-type name="Book">6</ref-type><contributors><authors><author>Higgins, J.P.T.; Green, S.</author></authors></contributors><titles><title>Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Chapter 8: Assessing Risk of Bias in Included Studies. The Cochrane Collaboration [updated March 2011].
</title></titles><dates><year>2011</year></dates><urls><related-urls><url>www.cochranehandbook.org</url></related-

urls></urls></record></Cite></EndNote>]. Sensitivity refers to the extent to which a study is likely to detect a true effect caused by exposure [ADDIN EN.CITE <EndNote><Cite><Author>Cooper</Author><Year>2016</Year><RecNum>32</RecNum><DisplayText>(Cooper et al., 2016)</DisplayText><record><rec-number>32</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509464913">32</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Cooper, G.</author><author>Lunn, R.</author><author>Agerstrand, M.</author><author>Glenn, B.</author><author>Kraft, A.</author><author>Luke, A.</author><author>Ratcliffe, J.</author></authors></contributors><titles><title>Study sensitivity: Evaluating the ability to detect effects in systematic reviews of chemical exposures</title><secondary-title>Environment International</secondary-title><alt-title>Environ Int</alt-title></titles><periodical><full-title>Environment International</full-title><abbr-1>Environ Int</abbr-1></periodical><alt-periodical><full-title>Environment International</full-title><abbr-1>Environ Int</abbr-1></alt-periodical><pages>605-610</pages><volume>92-93</volume><dates><year>2016</year></dates><isbn>ISSN 0160-4120
EISSN 1873-6750</isbn><label>3121908</label><urls><related-urls><url>http://dx.doi.org/10.1016/j.envint.2016.03.017</url></related-urls></urls><electronic-resource-num>10.1016/j.envint.2016.03.017</electronic-resource-num><language>English</language></record></Cite></EndNote>]. The general approach (described in this section) of study evaluation for epidemiology and animal studies is the same, but the specifics of applying the approach differ and thus they are described separately in the following sections.

Study evaluation considerations are specific to each study design, health effect, and agent. Subject-matter experts evaluate each group of studies to identify characteristics that bear on the informativeness of the results. For carcinogenicity, neurotoxicity, reproductive toxicity, and developmental toxicity, EPA guidance for study evaluation is available [ADDIN EN.CITE ADDIN EN.CITE.DATA].

Study evaluation for phthalates was conducted with two reviewers independently assessing each study, with inclusion of a pilot phase to assess and refine the evaluation process, comparison of decisions and reaching consensus among reviewers, and when necessary, resolution of differences by discussion among the reviewers, the chemical assessment team, or other technical experts. For studies that examined more than one endpoint or outcome, the evaluation process was performed separately for each outcome or endpoint, as the utility of a study may vary for different endpoints.

For each study[3] (specifically, an outcome or group of related outcomes in an individual study), in each evaluation domain, reviewers reached a consensus judgment of **Good**, **Adequate**, **Poor**, or **Critically Deficient**. It is important to stress that these evaluations were performed in the context of the study's utility for hazard identification of individual hazards. These terms are applied to each evaluation domain as follows:

- **Good** represents a judgment that there was appropriate study conduct relating to the domain, and any minor deficiencies that were noted would not be expected to influence the study results.
- **Adequate** indicates a judgment that there were experimental limitations relating to the domain, but that those limitations are not likely to be severe or to have a substantive impact on the results.
- **Poor** denotes identified biases or deficiencies that are interpreted as likely to have had a substantial impact on the results or that prevent reliable interpretation of the study findings.
- **Not reported** indicates that the information necessary to evaluate the domain question was not available in the study. Generally, this term carries the same functional interpretation as **Poor** for the purposes of the study confidence classification.
- **Critically Deficient** reflects a judgment that the experimental conduct relating to the domain question introduced a flaw so serious that the study should not be used without exceptional justification (e.g., it is the only study of its kind and may highlight possible research gaps). This judgment should only be used if there is an interpretation that the limitation(s) would be the primary driver of any observed effect(s), or if it makes the study uninterpretable.

Once the evaluation domains were considered, the identified strengths and limitations were combined to reach a study confidence classification of **High**, **Medium**, **Low**, or **Uninformative**. This classification was based on the reviewer judgments across the evaluation domains, and included consideration of the likely impact of the noted deficiencies in bias and sensitivity, or inadequate reporting, on the results. The classifications, which reflect a consensus judgment between reviewers, are defined as follows:

---

[3]Note: "study" is used instead of a more accurate term (e.g., "experiment") throughout these sections owing to an established familiarity within the field for discussing a study's risk of bias or sensitivity, etc. However, all evaluations discussed herein are explicitly conducted at the level of an individual outcome or group of outcomes within an (un)exposed group of animals or humans.

- **High Confidence**: No notable deficiencies or concerns were identified; the potential for bias is unlikely or minimal, and the study used sensitive methodology. In general, although classifications are not decided by "scoring," high confidence studies would reflect judgments of good across all or most evaluation domains.

- **Medium Confidence**: Possible deficiencies or concerns were noted, but the limitations are unlikely to be of a substantive degree. Generally, medium confidence studies will include adequate or good judgments across most domains, with the impact of any identified limitation not being judged as severe.

- **Low Confidence**: Deficiencies or concerns were noted, and the potential for substantive bias or inadequate sensitivity could have a significant impact on the study results or their interpretation. Typically, low confidence studies would have a poor evaluation for one or more domains (unless the impact of the particular limitations on the results is judged as unlikely to be severe).

- **Uninformative**: Serious flaw(s) make the study results unusable for informing hazard identification. Studies with critical deficiencies in any evaluation domain will almost always be classified as uninformative (see explanation above). Studies with multiple poor judgments across domains may also be considered uninformative, particularly when there is a robust database of studies on the outcome(s) of interest or when the impact of the limitations is viewed as severe.

Ratings are documented in Health Assessment Workspace Collaborative (HAWC), a free and open source web-based software application [4], for all animal studies and for epidemiology studies of most outcomes. Ratings of epidemiology studies for the remaining outcomes are documented in Microsoft Word.

4.1. Epidemiology study evaluation

Evaluation of epidemiology studies to assess bias and study sensitivity was conducted for the following domains: exposure measures, outcome measures, participant selection, potential confounding, analysis, selection of reported results, and study sensitivity (Table 3).

The principles and framework used for the evaluation of epidemiology studies are based on the Cochrane Risk of Bias in Non-randomized Studies (ROBINS) of interventions (ROBINS-I) [ADDIN EN.CITE <EndNote><Cite><Author>Sterne</Author><Year>2016</Year><RecNum>20</RecNum><DisplayText>(Sterne et al., 2016)</DisplayText><record><rec-number>20</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509462773">20</key></foreign-keys><ref-type name="Dataset">59</ref-type><contributors><authors><author>Sterne, J.</author><author>Higgins,

[4] Health Assessment Workspace Collaborative (HAWC): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals. [HYPERLINK "https://hawcproject.org/portal/"].

J.</author><author>Reeves, B.</author></authors></contributors><titles><title>ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions, Version 7 March 2016</title></titles><dates><year>2016</year></dates><pub-location>Ottawa, Canada</pub-location><publisher>Cochrane Methods Bias</publisher><label>3220127</label><work-type>Computer Program</work-type><urls><related-urls><url>http://www.riskofbias.info</url></related-urls></urls><language>English</language></record></Cite></EndNote>] but modified to address environmental and occupational exposures. The underlying philosophy of ROBINS-I is to describe attributes of an "ideal" study with respect to each of the evaluation domains (e.g., exposure measurement, outcome classification, etc.). Emphasis was placed on discerning a bias that would be expected to produce a substantive change in the estimated effect estimate. Core and prompting questions were used to collect information to guide evaluation of each domain. In addition, expected direction of bias is explicitly considered and the impact of a potential bias is incorporated into the decision-making process. Typical core and prompting questions used are presented in Table 4. Core questions are considered key concepts while prompting questions help the reviewer focus on relevant details under each key domain. Additional specific criteria for evaluating phthalate exposure measurement as well as phthalate-specific criteria on confounding and analysis are described below.

As discussed in the general evaluation methods, for each study, in each evaluation domain, reviewers reached a consensus on a value of **Good**, **Adequate**, **Poor**, or **Critically Deficient** based on risk of bias and sensitivity. Once the domains are classified, these ratings were combined to reach an overall study confidence classification of **High, Medium, Low,** or **Uninformative**. Once consensus was reached, the classifications were re-evaluated, looking at the variability within and between levels to ensure that the separation between the levels of confidence is appropriate and that no additional criteria need to be considered.

The reviewers also discussed additional data or analyses that could substantively change the evaluation or that would be needed to provide a meaningful interpretation of the results (e.g., different analyses that would allow direct comparison between studies). Study authors were contacted with requests for additional data or analyses that could substantively change the evaluation or allow for a more direct comparison of results across studies; study author responses were added to HERO.

Table [SEQ Table * ARABIC]. Domains of evaluation for epidemiology studies

Domain	Example information
Exposure measures	Source(s) of exposure (consumer products, occupational, an industrial accident) and source(s) of exposure data, blinding to outcome, level of detail for job history data, when measurements were taken, type of biomarker(s), assay information, reliability data from repeat measures studies, validation studies.

Domain	Example information
Outcome measures	Source of outcome (effect) measure, blinding to exposure status or level, how measured/classified, incident versus prevalent disease, evidence from validation studies, prevalence (or distribution summary statistics for continuous measures).
Participant selection	Study design, where and when was the study conducted, and who was included? Recruitment process, exclusion and inclusion criteria, type of controls, total eligible, comparison between participants and nonparticipants (or followed and not followed), final analysis group. Does the study include potential vulnerable/susceptible groups or lifestages?
Potential confounding	Background research on key confounders for specific populations or settings; participant characteristic data, by group; strategy/approach for consideration of potential confounding; strength of associations between exposure and potential confounders and between potential confounders and outcome; degree of exposure to the confounder in the population.
Analysis	Extent (and if applicable, treatment) of missing data for exposure, outcome, and confounders, approach to modeling, classification of exposure and outcome variables (continuous versus categorical), testing of assumptions, sample size for specific analyses, relevant sensitivity analyses.
Selective reporting	Are results presented with adequate detail for all of the endpoints of interest? Are results presented for the full sample as well as for specified subgroups? Were stratified analyses (effect modification) motivated by a specific hypothesis?
Sensitivity	What exposure range is spanned in this study? What are the ages of participants (e.g., not too young in studies of pubertal development)? What is the length of follow-up (for outcomes with long latency periods)? Choice of referent group and the level of exposure contrast between groups (i.e., the extent to which the "unexposed group" is truly unexposed, and the prevalence of exposure in the group designated as "exposed").

Table [SEQ Table * ARABIC]. Example question specification for evaluation of domains in epidemiology studies

Core question	Example prompting questions	Example follow-up questions
Exposure Does the exposure measure reliably distinguish between levels of exposure in a time window considered most relevant for a causal effect with respect to the development of the outcome?	For all: • Does the exposure measure capture the major source(s) of variability in exposure among the participants, considering intensity, frequency, and duration of exposure? • Does the exposure measure reflect a relevant time window? If not, can the relationship between measures in this time and the relevant time window be estimated reliably? • Was the exposure measurement likely to be affected by a knowledge of the outcome or by the presence of the outcome (i.e., reverse causality)? For case-control studies of occupational exposures: • Is exposure based on a comprehensive job history describing tasks, setting, time period, and use of specific materials? For biomarkers of exposure, general population: • Is a standard assay used? What are the intra- and inter-assay coefficients of variation? Is the assay likely to be affected by contamination? Are values less than the limit of detection dealt with adequately? • What exposure time-period is reflected by the biomarker? If the half-life is short, what is the correlation between serial measurements of exposure?	Is the degree of exposure misclassification likely to vary by exposure level? If the correlation between exposure measurements is moderate, is there an adequate statistical approach to ameliorate variability in measurements? If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?
Outcome Does the outcome measure reliably distinguish the presence or absence (or degree of severity) of the outcome?	For all: • Is disease ascertainment likely to be affected by knowledge of, or presence of, exposure (e.g., consider access to health care, if based on self-reported history of diagnosis)? For case-control studies: • Is the non-diseased comparison group (e.g., controls in a case-control study) based on objective criteria with little or no likelihood of inclusion of people with the disease? For mortality measures: • How well does cause of death data reflect occurrence of the disease in an individual? How well do mortality data reflect incidence of the disease? For diagnosis of disease measures: • Is diagnosis based on standard clinical criteria? If based on self-report of diagnosis, what is the validity of this measure?	Is there a concern that any outcome misclassification is non-differential, differential, or both? What is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?

Core question	Example prompting questions	Example follow-up questions
	For laboratory-based measures (e.g., hormone levels): Is a standard assay used? Does the assay have an acceptable level of inter-assay variability? Is the sensitivity of the assay appropriate for the outcome measure in this study population?	
Participant selection Is there evidence that selection into or out of the study (or analysis sample) was jointly related to exposure and to outcome?	For longitudinal cohort: • Did participants volunteer for the cohort based on knowledge of exposure and/or preclinical disease symptoms? Was entry into the cohort or continuation in the cohort related to exposure and outcome? For occupational cohort: • Did entry into the cohort begin with the start of the exposure? • Was follow-up or outcome assessment incomplete and if so, was follow-up related to both exposure and outcome status? • Could exposure produce symptoms that would result in a change in work assignment/work status ("healthy worker survivor effect")? For case-control study: • Were controls representative of population and time periods from which cases were drawn? • Are hospital controls selected from a group whose reason for admission is independent of exposure? • Could recruitment strategies, eligibility criteria, or participation rates result in differential participation relating to both disease and exposure? For population-based survey: Was recruitment based on advertisement to people with knowledge of exposure, outcome, and hypothesis?	Were differences in participant enrollment and follow-up evaluated to assess bias? If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)? Were appropriate analyses performed to address changing exposures over time in relation to symptoms? Is there a comparison of participants and non-participants to address whether or not differential selection is likely?
Confounding Is confounding of the effect of the exposure likely?	• Is confounding adequately addressed by considerations in… a. … participant selection (matching or restriction)? b. … accurate information on potential confounders, and statistical adjustment procedures? c. … lack of association between confounder and outcome, or confounder and exposure in the study? d. … information from other sources? Is the assessment of confounders based on a thoughtful review of published literature, potential relationships (e.g., as can be gained through directed acyclic graphing), minimizing potential over-control (e.g., inclusion of a variable on the pathway between exposure and outcome)?	If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?

Core question	Example prompting questions	Example follow-up questions
Analysis Does the analysis strategy and presentation convey the necessary familiarity with the data and assumptions?	• Are missing outcome, exposure, and covariate data recognized and, if necessary, accounted for in the analysis? • Does the analysis appropriately consider variable distributions and modeling assumptions? • Does the analysis appropriately consider subgroups of interest (e.g., based on variability in exposure level or duration, susceptible subgroups) • Is an appropriate analysis used for the study design? • Is effect modification considered, based on considerations developed a priori? • Does the study include additional analyses addressing potential biases or limitations (i.e., sensitivity analyses)?	If there is a concern about the potential for bias, what is the predicted direction or distortion of the bias on the effect estimate (if there is enough information)?

4.1.1. Evaluation of exposure

Biomarkers of Exposure in General Population Studies

Several factors were considered in the evaluation of biomarkers of phthalate exposure in general population studies (Table 5); confidence in the exposure measures was categorized accordingly based on specific factors or criteria. The evaluation of phthalate exposure was based on measured levels of metabolites—the primary monoesters for the shorter-chain phthalates (i.e., MBP for DBP; MIBP for DIBP, MBzP for BBP, MEP for DEP), and the secondary oxidative metabolites for the longer-chain phthalates (i.e., MEHHP, MEOHP, and MECPP for DEHP; MHiNP, MOiNP, and MCiOP for DINP) [ADDIN EN.CITE ADDIN EN.CITE.DATA].

An important consideration in the evaluation of biomarkers of phthalate exposure is the matrix from which phthalate metabolites are measured. Phthalate metabolite concentration in urine is considered to be the best proxy of exposure from all sources (ingested/absorbed/inhaled). One of the problems with phthalates measured in blood and other tissues is the potential for contamination from outside sources, especially during the collection and processing of samples [ADDIN EN.CITE <EndNote><Cite><Author>Calafat</Author><Year>2015</Year><RecNum>23</RecNum><DisplayText>(Calafat et al., 2015)</DisplayText><record><rec-number>23</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509462946">23</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Calafat, A. M.</author><author>Longnecker, M. P.</author><author>Koch, H. M.</author><author>Swan, S. H.</author><author>Hauser, R.</author><author>Goldman, L. R.</author><author>Lanphear, B. P.</author><author>Rudel, R. A.</author><author>Engel, S. M.</author><author>Teitelbaum, S. L.</author><author>Whyatt, R. M.</author><author>Wolff, M. S.</author></authors></contributors><titles><title>Optimal exposure biomarkers for

nonpersistent chemicals in environmental epidemiology</title><secondary-title>Environmental Health Perspectives</secondary-title><alt-title>Environ Health Perspect</alt-title></titles><periodical><full-title>Environmental Health Perspectives</full-title><abbr-1>Environ Health Perspect</abbr-1></periodical><alt-periodical><full-title>Environmental Health Perspectives</full-title><abbr-1>Environ Health Perspect</abbr-1></alt-periodical><pages>A166-A168</pages><volume>123</volume><number>7</number><dates><year>2015</year></dates><isbn>ISSN 0091-6765
EISSN 1552-9924</isbn><accession-num>26132373</accession-num><label>3045632</label><urls><related-urls><url>http://dx.doi.org/10.1289/ehp.1510041</url></related-urls></urls><electronic-resource-num>10.1289/ehp.1510041</electronic-resource-num><language>English</language></record></Cite></EndNote>]. Phthalate diesters present from exogenous contamination can be metabolized to the monoester metabolites by enzymes present in blood and other tissues (but not urine). Thus, metabolite measures in samples other than urine may be erroneously reflecting external phthalate sources. For these reasons, biomarker measures based on samples other than urine (e.g., serum, plasma, amniotic fluid, seminal fluid, amniotic fluid, breast milk) were considered to be *critically deficient* for all short-chain phthalates and for primary metabolites (e.g., MEHP, MINP) of long-chain phthalates. Samples other than urine can be used for secondary metabolites of long-chain phthalates as the oxidative metabolism required to break down primary metabolites does not exist in these samples. Cord blood, as a sample matrix, is considered *critically deficient* for all metabolites, since DEHP (and possibly DINP) containing plastics are widely used in medical settings, and thus, the concentrations of phthalates in cord blood may reflect exposure during delivery. In addition, studies that analyzed only phthalate diesters, rather than their metabolites, are considered *critically deficient* due to the potential for contamination.

Another consideration in the measurement of phthalate metabolite is the time of day. The half-life of phthalate metabolites is short, ranging from approximately 3 to <24 hours. The sources of exposure are numerous and widespread (e.g., cosmetics, food), and thus, exposures can occur frequently throughout the day. Variation in metabolite concentrations by time of day has been observed, and inclusion of the time of day of the sample collection into the analysis has been recommended for studies using spot urine samples [ADDIN EN.CITE <EndNote><Cite><Author>Johns</Author><Year>2015</Year><RecNum>24</RecNum><DisplayText>(Johns et al., 2015)</DisplayText><record><rec-number>24</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509463073">24</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Johns, L. E.</author><author>Ferguson, K. K.</author><author>Soldin, O. P.</author><author>Cantonwine, D. E.</author><author>Rivera-González, L. O.</author><author>Del Toro, L. V.</author><author>Calafat, A.

M.</author><author>Ye, X.</author><author>Alshawabkeh, A. N.</author><author>Cordero, J. F.</author><author>Meeker, J. D.</author></authors></contributors><titles><title>Urinary phthalate metabolites in relation to maternal serum thyroid and sex hormone levels during pregnancy: a longitudinal analysis</title><secondary-title>Reproductive Biology and Endocrinology</secondary-title><alt-title>Reprod Biol Endocrinol</alt-title></titles><periodical><full-title>Reproductive Biology and Endocrinology</full-title><abbr-1>Reprod Biol Endocrinol</abbr-1></periodical><alt-periodical><full-title>Reproductive Biology and Endocrinology</full-title><abbr-1>Reprod Biol Endocrinol</abbr-1></alt-periodical><pages>4</pages><volume>13</volume><number>1</number><dates><year>2015</year></dates><isbn>ISSN 1477-7827
EISSN 14777827</isbn><accession-num>25596636</accession-num><label>2804028</label><urls><related-urls><url>http://dx.doi.org/10.1186/1477-7827-13-4</url></related-urls></urls><electronic-resource-num>10.1186/1477-7827-13-4</electronic-resource-num><language>English</language></record></Cite></EndNote>]. An *ideal* study could include a validation component, allowing for the quantification of misclassification from using spot rather than 24-hour samples, or would incorporate the time of day of the sample collection into the analysis. More advanced statistical methods of adjusting for sampling conditions have been reported [ADDIN EN.CITE <EndNote><Cite><Author>Mortamais</Author><Year>2012</Year><RecNum>25</RecNum><DisplayText>(Mortamais et al., 2012)</DisplayText><record><rec-number>25</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509463147">25</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Mortamais, M.</author><author>Chevrier, C.</author><author>Philippat, C.</author><author>Petit, C.</author><author>Calafat, A. M.</author><author>Ye, X.</author><author>Silva, M. J.</author><author>Brambilla, C.</author><author>Eijkemans, M. J.</author><author>Charles, M. A.</author><author>Cordier, S.</author><author>Slama, R.</author></authors></contributors><titles><title>Correcting for the influence of sampling conditions on biomarkers of exposure to phenols and phthalates: a 2-step standardization method based on regression residuals</title><secondary-title>Environmental Health: A Global Access Science Source</secondary-title><alt-title>Environ Health</alt-title></titles><periodical><full-title>Environmental Health: A Global Access Science Source</full-title><abbr-1>Environ Health</abbr-1></periodical><alt-periodical><full-title>Environmental Health: A Global Access Science Source</full-title><abbr-1>Environ Health</abbr-1></alt-periodical><pages>29</pages><volume>11</volume><dates><year>2012</year></dates><isbn>ISSN 1476-069X
EISSN 1476-069X</isbn><accession-num>22537080</accession-num><label>1597770</label><urls><related-urls><url>http://dx.doi.org/10.1186/1476-069X-11-29</url></related-urls></urls><electronic-resource-num>10.1186/1476-069X-11-

29</electronic-resource-num><language>English</language></record></Cite></EndNote>], but are not currently commonly used. Further, while 24-hour samples account for the time of day, they may not be representative of typical exposure.

In addition to the time of day, another major concern with respect to interpretation of phthalate exposure measures is the reproducibility of measures over time (i.e., how well does a single sample reflect exposure or relative exposure for a given period of time). The short-term (1–12 weeks) reliability, measured by the intraclass correlation coefficient is approximately 0.3–0.6 for the shorter-chain metabolites (DBP, DIBP, BBP, and DEP) and 0.1–0.3 for the longer-chain metabolites (DEHP and DINP) [ADDIN EN.CITE <EndNote><Cite><Author>Johns</Author><Year>2015</Year><RecNum>24</RecNum><DisplayText>(Johns et al., 2015)</DisplayText><record><rec-number>24</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509463073">24</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Johns, L. E.</author><author>Ferguson, K. K.</author><author>Soldin, O. P.</author><author>Cantonwine, D. E.</author><author>Rivera-González, L. O.</author><author>Del Toro, L. V.</author><author>Calafat, A. M.</author><author>Ye, X.</author><author>Alshawabkeh, A. N.</author><author>Cordero, J. F.</author><author>Meeker, J. D.</author></authors></contributors><titles><title>Urinary phthalate metabolites in relation to maternal serum thyroid and sex hormone levels during pregnancy: a longitudinal analysis</title><secondary-title>Reproductive Biology and Endocrinology</secondary-title><alt-title>Reprod Biol Endocrinol</alt-title></titles><periodical><full-title>Reproductive Biology and Endocrinology</full-title><abbr-1>Reprod Biol Endocrinol</abbr-1></periodical><alt-periodical><full-title>Reproductive Biology and Endocrinology</full-title><abbr-1>Reprod Biol Endocrinol</abbr-1></alt-periodical><pages>4</pages><volume>13</volume><number>1</number><dates><year>2015</year></dates><isbn>ISSN 1477-7827
EISSN 14777827</isbn><accession-num>25596636</accession-num><label>2804028</label><urls><related-urls><url>http://dx.doi.org/10.1186/1477-7827-13-4</url></related-urls></urls><electronic-resource-num>10.1186/1477-7827-13-4</electronic-resource-num><language>English</language></record></Cite></EndNote>]. The sensitivity of a single urine sample to ranking individuals into categories (e.g., above and below the median, or by tertile of exposure) has been shown to be between approximately 0.5 and 0.75 [ADDIN EN.CITE ADDIN EN.CITE.DATA]. Thus, the use of a single sample would be expected to introduce nondifferential exposure misclassification into the analysis. More samples are needed to measure exposure to the longer-chain phthalates with sufficient precision. Multiple spot urine samples at different times of day or pooled samples from several days are preferred to 24-hour voids over a single day [ADDIN EN.CITE ADDIN EN.CITE.DATA].

Other issues considered in the evaluation of exposure measures in general population studies included the discussion of quality control procedures used in laboratory analyses (although many studies lack details on this issue), and the proportion of samples that were above the level of detection. The exposure levels (expressed as a median or geometric mean) or range of exposure levels encompassed by the study population was also considered. Differing descriptive measures of exposure levels (median, geometric mean, or range) affect the ability to compare results across studies. In addition, a limited exposure range could limit the ability of a study to detect an effect of exposure on a given endpoint [ADDIN EN.CITE <EndNote><Cite><Author>Cooper</Author><Year>2016</Year><RecNum>32</RecNum><DisplayText>(Cooper et al., 2016)</DisplayText><record><rec-number>32</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509464913">32</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Cooper, G.</author><author>Lunn, R.</author><author>Agerstrand, M.</author><author>Glenn, B.</author><author>Kraft, A.</author><author>Luke, A.</author><author>Ratcliffe, J.</author></authors></contributors><titles><title>Study sensitivity: Evaluating the ability to detect effects in systematic reviews of chemical exposures</title><secondary-title>Environment International</secondary-title><alt-title>Environ Int</alt-title></titles><periodical><full-title>Environment International</full-title><abbr-1>Environ Int</abbr-1></periodical><alt-periodical><full-title>Environment International</full-title><abbr-1>Environ Int</abbr-1></alt-periodical><pages>605-610</pages><volume>92-93</volume><dates><year>2016</year></dates><isbn>ISSN 0160-4120
EISSN 1873-6750</isbn><label>3121908</label><urls><related-urls><url>http://dx.doi.org/10.1016/j.envint.2016.03.017</url></related-urls></urls><electronic-resource-num>10.1016/j.envint.2016.03.017</electronic-resource-num><language>English</language></record></Cite></EndNote>].

Other aspects of exposure measures were considered within the categories of confounding (e.g., the potential correlations among phthalate metabolites) or analysis (e.g., methods used to address urine dilution). These considerations are discussed further in subsequent sections of confounding and analysis, as applicable.

Table [SEQ Table * ARABIC]. Evaluation of exposure biomarkers in general population studies of phthalates

Criteria			
Level		**Short-chain (DEP, DBP, DIBP, BBP)**	**Long-chain (DEHP, DINP)**
Good		▪ Two or more urine samples within the etiologically relevant period (±1–3 mo) **and** ▪ High proportion (>50%) above the LOD **and** ▪ Discussion of laboratory QC procedures **or** no discussion of laboratory QC procedures but analysis by an experienced laboratory (e.g., Centers for Disease Control and Prevention [CDC])	▪ Three or more urine samples within etiologically relevant time period (±1–3 mo) and analysis includes a summed variable, or similar results seen with each of the metabolites from the parent compound **and** ▪ High proportion (>50%) above the LOD **and** ▪ Discussion of laboratory QC procedures or no discussion of laboratory QC procedures but analysis by an experienced laboratory (e.g., CDC)
Adequate		▪ One urine sample within etiologically relevant period (±1–3 mo) **and** ▪ High proportion (>50%) above the LOD	▪ Two or more urine samples within etiologically relevant period (±1–3 mo) and analysis includes a summed variable, or similar results seen with each of the metabolites from the parent compound **and** ▪ High proportion (>50%) above the LOD
Poor		▪ One urine sample; sample collection may be 3–24 mo outside etiologically relevant period **and** ▪ High proportion (>50%) above the LOD	▪ One urine sample within etiologically relevant period (±1–3 mo) and high proportion (>50%) above the LOD **or** ▪ Does not include summed variable (e.g., ∑DEHP) and results differ among the metabolites (e.g., MEHP, MEOHP, and MEHHP have different results)
Critically Deficient		▪ Biomarker measured in tissue other than urine **or** ▪ Measures in urine likely to be affected by differential misclassification (e.g., after disease diagnosis) **or** ▪ Low proportion (<50%) above the LOD	▪ Biomarker measured in tissue other than urine if only primary metabolites (MEHP, MINP are analyzed) **or** ▪ Biomarker measured in cord blood for all metabolites **or** ▪ Measures in urine likely to be affected by differential misclassification (e.g., after disease diagnosis) **or** ▪ Low proportion (<50%) above the LOD

Studies of Occupational Exposures

Exposure to phthalates can occur in many different occupational settings, including the production of products containing polyvinyl chloride and rubber, pharmaceuticals, and cosmetics.

However, without the use of sensitive and specific biomarkers of exposure as described for general population studies above, it is difficult to determine the specific phthalates and level of exposure an individual is subject to. [HYPERLINK \l "_ENREF_21" \o "Hines, 2009 #33"] examined exposure based on mid- and end-shift urine samples in workers in eight different settings (phthalate manufacturing, polyvinyl chloride [PVC] film manufacturing, PVC compounding, production of vehicle filters, rubber boots, rubber hoses, or rubber gaskets, and nail-only salons). Monoester metabolites of DEHP, DBP, DEP, BBP, DIBP, dimethyl phthalate (DMP), or DnOP were analyzed in urine samples. Visits to each worksite examined the process and materials used in different locations. In most cases, increases (mid- to end-shift, or end-shift compared to National Health and Examination Survey [NHANES] general population data) were seen for the metabolites corresponding to the parent compounds that were said to be used at the site. In some places, however, increases in metabolites were seen for phthalates that had not been identified. Other studies within a specific plant(s) (e.g., a PVC pellet plant) observed considerable variation in exposure based either on air monitoring or individual worker urine samples [ADDIN EN.CITE ADDIN EN.CITE.DATA]. Thus, for evaluation of the exposure measures in an occupational setting, the extent and use of individual-level exposure data was considered (Table 6); studies that used personal samples were classified with higher confidence.

Several studies were identified that used a job exposure matrix covering 348 jobs developed by [HYPERLINK \l "_ENREF_49" \o "Van Tongeren, 2002 #36"] to characterize occupational exposure to endocrine-disrupting chemicals in population-based studies (e.g., case-control studies or registry-based studies). Phthalates were one of seven types of chemicals included in this work (in addition to pesticides, polychlorinated organic compounds, alkylphenolic compounds, bi-phenolic compounds, heavy metals, and an "other" category). However, data to support the reliability and validity of this job exposure matrix for characterizing exposure to phthalates (as a general group, or for individual phthalates) were not available. [HYPERLINK \l "_ENREF_50" \o "Vrijheid, 2003 #37"] noted that phthalates were one of two categories with a great degree of disagreement (i.e., low inter-rater correlation) among industrial hygienists estimating exposure. Another study examined urinary DEHP metabolite levels in pregnant women who were working; urinary metabolite levels in the 11 women classified as "possible occupational phthalate exposure" based on the job exposure matrix were lower than those of the 103 women classified as "unlikely phthalate exposure" (∑DEHP median 320 and 645 nmol/L, respectively, in the possible and unlikely groups, $p = 0.34$) [ADDIN EN.CITE <EndNote><Cite><Author>Chevrier</Author><Year>2012</Year><RecNum>38</RecNum><DisplayText>(Chevrier et al., 2012)</DisplayText><record><rec-number>38</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509465382">38</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Chevrier, C.</author><author>Petit, C.</author><author>Philippat, C.</author><author>Mortamais, M.</author><author>Slama,

R.</author><author>Rouget, F.</author><author>Calafat, A. M.</author><author>Ye, X.</author><author>Silva, M. J.</author><author>Charles, M. A.</author><author>Cordier, S.</author></authors></contributors><titles><title>Maternal urinary phthalates and phenols and male genital anomalies</title><secondary-title>Epidemiology</secondary-title><alt-title>Epidemiology</alt-title></titles><periodical><full-title>Epidemiology</full-title><abbr-1>Epidemiology</abbr-1></periodical><alt-periodical><full-title>Epidemiology</full-title><abbr-1>Epidemiology</abbr-1></alt-periodical><pages>353-356</pages><volume>23</volume><number>2</number><dates><year>2012</year></dates><isbn>ISSN 1044-3983
EISSN 1531-5487</isbn><accession-num>22317818</accession-num><label>1597808</label><work-type>Letter</work-type><urls><related-urls><url>http://dx.doi.org/10.1097/EDE.0b013e318246073e</url></related-urls></urls><electronic-resource-num>10.1097/EDE.0b013e318246073e</electronic-resource-num><language>English</language></record></Cite></EndNote>]. Based on these studies, use of job exposure matrices to assess occupational exposure to phthalates (i.e., job exposure matrices) was considered to be *critically deficient*; this classification can be revisited if new data pertaining to the reliability and validity of the method become available. Additional validation would allow a population-based occupational study to be classified as *adequate* for exposure measurement, but no higher (Table 6).

Table [SEQ Table * ARABIC]. Evaluation of exposure without biomarkers in occupational studies of phthalates

Level		Criteria
Good		**For a specific job site(s):** * Current exposure based on personal samples (urine or air) covering at least one full shift. Ideally, this would cover all workers or randomly selected workers within specific areas/jobs/tasks, allowing for examination of variation in exposure among workers at a particular worksite, but this is not required. * (Where applicable): Past exposures based on samples collected spanning time period and variations at work locations, tasks, and conditions (i.e., some kind of validation done for estimation of past exposures).
Adequate		* **For a specific job site(s):** with known exposure to specific phthalates, among workers with likely exposure, current exposure based on this work (rather than metabolite levels) can be used for comparison with background (general population) if there is reason to believe exposures at this site are higher than background levels. * **For population-based occupational studies:** job exposure matrix that incorporates industry, time period, tasks, and material used, and has validation data confirming its ability to differentiate between exposure levels (e.g., using urine biomonitoring from a subset or previous research). * (Where applicable): Past exposures based on historical data, with limited details.

Level		Criteria
Poor		**For a specific job site(s):** with known exposure to specific phthalates, without information on proportion of workers that are likely to be exposed (this may be an insensitive exposure measure).
Critically Deficient		**For population-based occupational studies:** * Self-report of "exposed to phthalates" or * Job exposure without validation data or with data indicating that it cannot differentiate between exposure levels.

4.1.2. Evaluation of confounding

Evaluation of the adequacy of a study's approach to confounding was not based on a checklist of variables that are either included or not included in a model. Rather, it considered the strategy used for identifying confounders, recognizing that not all "risk factors" are confounders and the role of a variable as a potential mediator (intermediary on causal pathway) or confounder [ADDIN EN.CITE <EndNote><Cite><Author>Kaufman</Author><Year>2004</Year><RecNum>39</RecNum><DisplayText>(Kaufman et al., 2004)</DisplayText><record><rec-number>39</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509465460">39</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Kaufman, J. S.</author><author>MacLehose, R. F.</author><author>Kaufman, S.</author></authors></contributors><titles><title>A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation</title><secondary-title>Epidemiologic Perspectives and Innovations</secondary-title><alt-title>Epidemiologic Perspectives and Innovations</alt-title></titles><periodical><full-title>Epidemiologic Perspectives and Innovations</full-title><abbr-1>Epidemiologic Perspectives and Innovations</abbr-1></periodical><alt-periodical><full-title>Epidemiologic Perspectives and Innovations</full-title><abbr-1>Epidemiologic Perspectives and Innovations</abbr-1></alt-periodical><pages>4</pages><volume>1</volume><number>1</number><dates><year>2004</year></dates><isbn>ISSN 1742-5573
EISSN 17425573</isbn><accession-num>15507130</accession-num><label>3090081</label><urls><related-urls><url>http://dx.doi.org/10.1186/1742-5573-1-4</url></related-urls></urls><electronic-resource-num>10.1186/1742-5573-1-4</electronic-resource-num><language>English</language></record></Cite></EndNote>]. In certain cases the impact of confounding could be relatively small [ADDIN EN.CITE <EndNote><Cite><Author>Blair</Author><Year>2007</Year><RecNum>40</RecNum><DisplayText>(Blair et al., 2007)</DisplayText><record><rec-number>40</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps"

timestamp="1509465528">40</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Blair, A.</author><author>Stewart, P.</author><author>Lubin, J. H.</author><author>Forastiere, F.</author></authors></contributors><titles><title>Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures</title><secondary-title>American Journal of Industrial Medicine</secondary-title><alt-title>Am J Ind Med</alt-title></titles><periodical><full-title>American Journal of Industrial Medicine</full-title><abbr-1>Am J Ind Med</abbr-1></periodical><alt-periodical><full-title>American Journal of Industrial Medicine</full-title><abbr-1>Am J Ind Med</abbr-1></alt-periodical><pages>199-207</pages><volume>50</volume><number>3</number><dates><year>2007</year></dates><isbn>ISSN 0271-3586
EISSN 1097-0274</isbn><accession-num>17096363</accession-num><label>729541</label><work-type>Review</work-type><urls><related-urls><url>http://dx.doi.org/10.1002/ajim.20281</url></related-urls></urls><electronic-resource-num>10.1002/ajim.20281</electronic-resource-num><language>English</language></record></Cite></EndNote>], but nevertheless, it needs to be evaluated. An ideal study would include a justification for inclusion or exclusion of variables based on biological considerations and the associations observed within the study population. For more recent studies, it might also include a description of causal structure through, for example, the use of directed acyclic graphs. Other considerations include the descriptive information provided regarding potential confounders, including how these variables vary by outcome status or levels of exposure, and results that allow the reader to observe the impact of inclusion of different variables. It is also important to note that addressing relevant confounding factors can occur at the population selection step through, for example, matching or restriction.

For each outcome, the evaluation strategy includes a specific list of potential variables that are known to be relatively strong predictors of the outcome. These may or may not be relevant confounders for phthalate exposure. Potential confounding, assessed by degree of co-occurrence, among phthalates is also a consideration. Since some different phthalates may be used in similar applications, potential confounding among phthalates is an important consideration. Several of the phthalates have moderate correlations with each other. When results are similar for two moderately (or higher) correlated phthalates in a study, confounding by other phthalate exposures should be considered as a possible explanation. An ideal study would have this accounted for in its analysis. Secondary metabolites of DEHP (and likely other long-chain phthalates) are highly correlated amongst each other [ADDIN EN.CITE <EndNote><Cite><Author>Johns</Author><Year>2015</Year><RecNum>24</RecNum><DisplayText>(Johns et al., 2015)</DisplayText><record><rec-number>24</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509463073">24</key></foreign-keys><ref-type name="Journal Article">17</ref-

type><contributors><authors><author>Johns, L. E.</author><author>Ferguson, K. K.</author><author>Soldin, O. P.</author><author>Cantonwine, D. E.</author><author>Rivera-González, L. O.</author><author>Del Toro, L. V.</author><author>Calafat, A. M.</author><author>Ye, X.</author><author>Alshawabkeh, A. N.</author><author>Cordero, J. F.</author><author>Meeker, J. D.</author></authors></contributors><titles><title>Urinary phthalate metabolites in relation to maternal serum thyroid and sex hormone levels during pregnancy: a longitudinal analysis</title><secondary-title>Reproductive Biology and Endocrinology</secondary-title><alt-title>Reprod Biol Endocrinol</alt-title></titles><periodical><full-title>Reproductive Biology and Endocrinology</full-title><abbr-1>Reprod Biol Endocrinol</abbr-1></periodical><alt-periodical><full-title>Reproductive Biology and Endocrinology</full-title><abbr-1>Reprod Biol Endocrinol</abbr-1></alt-periodical><pages>4</pages><volume>13</volume><number>1</number><dates><year>2015</year></dates><isbn>ISSN 1477-7827
EISSN 14777827</isbn><accession-num>25596636</accession-num><label>2804028</label><urls><related-urls><url>http://dx.doi.org/10.1186/1477-7827-13-4</url></related-urls></urls><electronic-resource-num>10.1186/1477-7827-13-4</electronic-resource-num><language>English</language></record></Cite></EndNote>]. It is not necessary, and could introduce bias, to adjust for multiple metabolites from the same phthalate.

Details of the evaluation criteria for confounding can be found in Table 7.

Table [SEQ Table * ARABIC]. Criteria for evaluation of confounding

Level		Criteria
Good		• Conveys thoughtful strategy for identifying confounders. This may include justification for inclusion or exclusion of variables (based on a priori biological considerations, statistical analysis, or results in the published literature) with the recognition that not all "risk factors" are confounders. • Inclusion in model not based solely on statistical significance criteria (e.g., p<0.05 from stepwise regression) • Does not include variables on the causal pathway (i.e., intermediaries). • Shows progression of adjustment with different models and presents other data relevant to potential for confounding (e.g., distribution of variables by exposure category or discussion of likelihood that residual confounding could explain the observed effect). • Descriptive information for relevant population characteristics/potential confounders presented (with amount of missing data noted).
Adequate		• Conveys some discussion of strategy for identifying confounders. OR • Shows progression of adjustment with different models and presents other data relevant to potential for confounding (e.g., distribution of variables by exposure category or discussion of likelihood that residual confounding could explain the observed effect). AND all of the following:

		• Inclusion in model not based solely on statistical significance criteria (e.g., p<0.05 from stepwise regression) • Does not include variables on the causal pathway (i.e., intermediaries). • Descriptive information for relevant population characteristics/potential confounders presented (with amount of missing data noted).
Poor		• Strategy for evaluating confounding is unclear or is not recommended (e.g., based on statistical significance criteria only). OR • Descriptive information on population characteristics or potential confounders not presented. OR • Some residual confounding is likely, given the observed effect and likely measurement misclassification for a confounder present in the study.
Critically deficient		• Established intermediary included as a confounder. • Confounding present and not accounted for (i.e., the relevant variable is associated with the outcome and exposure in the study and was not accounted for in the analysis), and lack of consideration of the confounder could explain the results observed with the exposure.

4.1.3. Evaluation of analysis

The assessment of confounding is just one component of the overall analysis of epidemiology studies. The analytic strategy and attention to detail can lead to varying levels of confidence in the results presented and could overestimate or underestimate the observed association between the exposure of interest and the outcome. An ideal study would convey a thoughtful and thorough description of the analytical approach, and descriptive data for key variables (e.g., exposure measures, outcome measures), including the amount of missing data (or proportion less than the limit of detection [LOD]). The ideal analysis would use an appropriate and well thought out modeling approach for the study design (e.g., logistic regression for case-control data) and specify the covariates used in the final model; the methods should be described in enough detail such that they could be applied to the data from another study. In addition, the results should be presented with sufficient detail to enable estimation of effect estimates and precision of the estimates (e.g., standard error [SE] or confidence interval [CI]).

The development of models could include consideration of additional variables for purposes other than to address confounding (i.e., a strong predictor could be included in a model to improve precision of estimates). A thoughtful strategy for modeling should take into account the known literature on relationships between all of the potential model variables under consideration and would recognize the potential for controlling for intermediate variables, the inducement of unexpected relationships that arise conditionally upon the statistical control of other covariates, and the potential for diminished statistical power from over-control of non-predictive covariates. In addition, statistical modeling assumptions and model diagnostics should be presented. Other analysis considerations include exploration of various shapes of exposure-responses, effect

modification with a priori rationale for stratification variables and sufficient numbers in subgroups for analysis, relevant sensitivity analyses addressing issues of outcome classification, selection, different methods for adjusting for urine dilution, and other features of the design or data. Through these methods, the authors should convey a thorough exploration of the data and address the robustness of the results (i.e., to what extent the results depend on specific decisions made regarding the analysis).

Adjustment in the measurement of urinary biomarkers is used to account for dilution-dependent sample variation in urine concentrations at the time of measurement and to control for measurement error bias caused by dilution of the urine specimen. Several approaches are used to standardize exposure (e.g., creatinine, osmolality, or specific gravity) and methods of analysis (e.g., various ways to include the dilution variable in model). Most investigators agree that an adjustment in the measurement of a urinary biomarker is needed, but there is controversy over which adjustment is best. Ideally, a study would use two or more different methods and compare results among them to assess impact on interpretation of results, as well as consider the underlying causal relationship between exposure and outcome [ADDIN EN.CITE <EndNote><Cite><Author>O'Brien</Author><Year>2016</Year><RecNum>41</RecNum><DisplayText>(O'Brien et al., 2016)</DisplayText><record><rec-number>41</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509465689">41</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>O'Brien, K. M.</author><author>Upson, K.</author><author>Cook, N. R.</author><author>Weinberg, C. R.</author></authors></contributors><titles><title>Environmental chemicals in urine and blood: Improving methods for creatinine and lipid adjustment</title><secondary-title>Environmental Health Perspectives</secondary-title><alt-title>Environ Health Perspect</alt-title></titles><periodical><full-title>Environmental Health Perspectives</full-title><abbr-1>Environ Health Perspect</abbr-1></periodical><alt-periodical><full-title>Environmental Health Perspectives</full-title><abbr-1>Environ Health Perspect</abbr-1></alt-periodical><pages>220-227</pages><volume>124</volume><number>2</number><dates><year>2016</year></dates><isbn>ISSN 0091-6765
EISSN 1552-9924</isbn><accession-num>26219104</accession-num><label>3771537</label><urls><related-urls><url>http://dx.doi.org/10.1289/ehp.1509693</url></related-urls></urls><electronic-resource-num>10.1289/ehp.1509693</electronic-resource-num><language>English</language></record></Cite></EndNote>]. Alternatively, including either specific gravity or osmolality as measures of urine dilution in the model may be preferred [ADDIN EN.CITE <EndNote><Cite><Author>Johns</Author><Year>2015</Year><RecNum>24</RecNum><DisplayText>(Johns et al., 2015)</DisplayText><record><rec-number>24</rec-number><foreign-

keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509463073">24</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Johns, L. E.</author><author>Ferguson, K. K.</author><author>Soldin, O. P.</author><author>Cantonwine, D. E.</author><author>Rivera-González, L. O.</author><author>Del Toro, L. V.</author><author>Calafat, A. M.</author><author>Ye, X.</author><author>Alshawabkeh, A. N.</author><author>Cordero, J. F.</author><author>Meeker, J. D.</author></authors></contributors><titles><title>Urinary phthalate metabolites in relation to maternal serum thyroid and sex hormone levels during pregnancy: a longitudinal analysis</title><secondary-title>Reproductive Biology and Endocrinology</secondary-title><alt-title>Reprod Biol Endocrinol</alt-title></titles><periodical><full-title>Reproductive Biology and Endocrinology</full-title><abbr-1>Reprod Biol Endocrinol</abbr-1></periodical><alt-periodical><full-title>Reproductive Biology and Endocrinology</full-title><abbr-1>Reprod Biol Endocrinol</abbr-1></alt-periodical><pages>4</pages><volume>13</volume><number>1</number><dates><year>2015</year></dates><isbn>ISSN 1477-7827
EISSN 14777827</isbn><accession-num>25596636</accession-num><label>2804028</label><urls><related-urls><url>http://dx.doi.org/10.1186/1477-7827-13-4</url></related-urls></urls><electronic-resource-num>10.1186/1477-7827-13-4</electronic-resource-num><language>English</language></record></Cite></EndNote>]. For purposes of study quality evaluation, no adjustment or consideration of urine dilution in a study is considered a study limitation, but this limitation is expected to have only a small impact. For phthalates, only two studies were identified where urine dilution adjustments changed the interpretation of the results: one that had a very small sample size [ADDIN EN.CITE <EndNote><Cite><Author>Meeker</Author><Year>2009</Year><RecNum>42</RecNum><DisplayText>(Meeker et al., 2009)</DisplayText><record><rec-number>42</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509465759">42</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Meeker, J. D.</author><author>Hu, H.</author><author>Cantonwine, D. E.</author><author>Lamadrid-Figueroa, H.</author><author>Calafat, A. M.</author><author>Ettinger, A. S.</author><author>Hernandez-Avila, M.</author><author>Loch-Caruso, R.</author><author>Tellez-Rojo, M. M.</author></authors></contributors><titles><title>Urinary phthalate metabolites in relation to preterm birth in Mexico city</title><secondary-title>Environmental Health Perspectives</secondary-title><alt-title>Environ Health Perspect</alt-title></titles><periodical><full-title>Environmental Health Perspectives</full-title><abbr-1>Environ Health Perspect</abbr-1></periodical><alt-periodical><full-title>Environmental Health Perspectives</full-title><abbr-1>Environ Health Perspect</abbr-1></alt-periodical><pages>1587-

1592</pages><volume>117</volume><number>10</number><dates><year>2009</year></dates><isbn>ISSN 0091-6765
EISSN 1552-9924</isbn><accession-num>20019910</accession-num><label>673483</label><urls><related-urls><url>http://dx.doi.org/10.1289/ehp.0800522</url></related-urls></urls><electronic-resource-num>10.1289/ehp.0800522</electronic-resource-num><language>English</language></record></Cite></EndNote>] and one that focused on children [ADDIN EN.CITE <EndNote><Cite><Author>Boas</Author><Year>2010</Year><RecNum>43</RecNum><DisplayText>(Boas et al., 2010)</DisplayText><record><rec-number>43</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509465809">43</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Boas, M.</author><author>Frederiksen, H.</author><author>Feldt-Rasmussen, U.</author><author>Skakkebaek, N. E.</author><author>Hegedus, L.</author><author>Hilsted, L.</author><author>Juul, A.</author><author>Main, K. M.</author></authors></contributors><titles><title>Childhood exposure to phthalates: Associations with thyroid function, insulin-like growth factor I, and growth</title><secondary-title>Environmental Health Perspectives</secondary-title><alt-title>Environ Health Perspect</alt-title></titles><periodical><full-title>Environmental Health Perspectives</full-title><abbr-1>Environ Health Perspect</abbr-1></periodical><alt-periodical><full-title>Environmental Health Perspectives</full-title><abbr-1>Environ Health Perspect</abbr-1></alt-periodical><pages>1458-1464</pages><volume>118</volume><number>10</number><dates><year>2010</year></dates><isbn>ISSN 0091-6765
EISSN 1552-9924</isbn><accession-num>20621847</accession-num><label>673235</label><urls><related-urls><url>http://dx.doi.org/10.1289/ehp.0901331</url></related-urls></urls><electronic-resource-num>10.1289/ehp.0901331</electronic-resource-num><language>English</language></record></Cite></EndNote>]. This adjustment may be a bigger concern for studies of children, as creatinine is strongly correlated with sex, age, and body measures in children [ADDIN EN.CITE <EndNote><Cite><Author>Skinner</Author><Year>1996</Year><RecNum>44</RecNum><DisplayText>(Skinner et al., 1996)</DisplayText><record><rec-number>44</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509465863">44</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Skinner, A. M.</author><author>Addison, G. M.</author><author>Price, D. A.</author></authors></contributors><titles><title>Changes in the urinary excretion of creatinine, albumin and N-acetyl-beta-D-glucosaminidase with increasing age and maturity in healthy schoolchildren</title><secondary-title>European Journal of Pediatrics</secondary-title><alt-title>Eur J Pediatr</alt-title></titles><periodical><full-

title>European Journal of Pediatrics</full-title><abbr-1>Eur J Pediatr</abbr-1></periodical><alt-periodical><full-title>European Journal of Pediatrics</full-title><abbr-1>Eur J Pediatr</abbr-1></alt-periodical><pages>596-602</pages><volume>155</volume><number>7</number><dates><year>1996</year></dates><isbn>ISSN 0340-6199
EISSN 1432-1076</isbn><accession-num>8831085</accession-num><label>2736265</label><urls><related-urls><url>http://dx.doi.org/10.1007/BF01957912</url></related-urls></urls><electronic-resource-num>10.1007/BF01957912</electronic-resource-num><language>English</language></record></Cite></EndNote>].

Details of the evaluation criteria for analysis can be found in Table 8.

Table [SEQ Table * ARABIC]. Criteria for evaluation of analysis

Level		Criteria
Good		• Appropriate analysis methods used. • Quantitative results presented (e.g., effect estimates and confidence limits). • Descriptive information about outcome and exposure provided (where applicable): ○ Amount of missing data noted and addressed appropriately. ○ For exposure, includes LOD (and percentage less than LOD) and discussion of cut-points and transformations. • Includes analyses that address robustness of findings (e.g., examination of shape of exposure-response, relevant sensitivity analyses). Effect modification only examined with *a priori* rationale and sufficient numbers.
Adequate		Same as Good, except: • Descriptive information about exposure provided, but may be incomplete; might not have discussed missing data, or cut-points, or shape of distribution. OR • Some important analyses that address the robustness of findings are not performed. •
Poor		• Descriptive information about exposure levels not provided (where applicable). OR • Effect estimate presented without standard error or confidence interval. OR • Non-optimal analysis methods used (e.g., correlation instead of linear regression)
Critically deficient		• Results presented as statistically "significant"/ "not significant" or just p-values (i.e., without including effect estimates). OR

		• Effect modification examined without clear *a priori* rationale and without providing main effects. OR Analysis methods were not appropriate for the design or data.

4.2. Animal study evaluation

Evaluation of animal studies to assess risk of bias and study sensitivity was conducted for the following domains: reporting quality, selection or performance bias, confounding/variable control, reporting or attrition bias, exposure methods sensitivity, and outcome measures and results display (Table 9). Human relevance of the animal model or endpoint is not addressed during study evaluation; rather, those considerations are addressed as either part of the PECO (known non-relevant models or endpoints should be excluded during screening) or during evidence synthesis (discussed later).

As discussed in the general evaluation methods, for each study, in each evaluation domain, reviewers reached a consensus on a value of **Good**, **Adequate**, **Poor**, or **Critically Deficient** based on risk of bias and sensitivity. Once the domains are classified, these ratings were combined to reach an overall study confidence classification of **High, Medium, Low,** or **Uninformative**. Once consensus was reached, the classifications were re-evaluated, looking at the variability within and between levels to ensure that the separation between the levels of confidence is appropriate and that no additional criteria need to be considered.

Table [SEQ Table * ARABIC]. General criteria to evaluate outcomes from animal toxicology studies

Domain and Metric	*Criteria*
Reporting Quality Reporting of information necessary for study evaluation	• Key information necessary for study evaluation (study would be deemed critically deficient if not reported[1]): Species; sex; test article description; levels and duration of exposure; route of administration; endpoints investigated; qualitative or quantitative results. Important information, which should also be reported, is listed below. The brackets contain secondary information that would ideally be reported and, based on the needs of a given assessment, may be considered important, or key, information. • *Test animal* – strain; source (e.g., vendor); husbandry procedures (e.g., housing, feed, mating); [baseline health (e.g., colony monitoring procedures); age or body weight at start of study]. • *Exposure methods* – test article source; description of vehicle control; methods of administration (e.g., gavage volume, exposure chamber); [information on stability; purity; analytical verification methods]. • *Experimental design* – periodicity of exposure; animal age/life-stage during exposure and at endpoint evaluation(s); [timing of endpoint evaluation(s) (e.g., latency between exposure and testing)]. • *Endpoint evaluations* – procedural details to understand how endpoints were measured; procedural controls, including information on positive and negative controls; [related details (e.g., biological matrix or specific region of tissue/ organ evaluated); information on other manipulations (e.g., surgery, co-treatment)]. • *Results presentation* – presents findings for all endpoints of interest that were investigated; information on variability; experimental units assessed; sample size; statistical procedures; (related details- e.g., maternal toxicity in developmental studies; handling of early mortality in long-term bioassays). [1] Although such decisions should be made on an assessment-specific basis, if this information is not reported, it is generally not useful to reach out to study authors. However, for other missing study details that might change study confidence conclusions if it they were available, efforts should be made to reach out to study authors.
Selection or Performance Bias Allocation of animals to experimental groups	Ideally, animal studies are randomized, with each animal or litter having an equal chance of being assigned to any experimental group, including controls, and allocation procedures sufficiently described. Less ideally, but generally adequate or good, are studies indicating normalization of experimental groups prior to exposure, for example according to body weight or litter, but without indication of randomization. The least preferred situation is studies with no indication of how groups were assigned.
Blinding of investigators, particularly during outcome assessment	In a good study, observational bias can be reduced through concealment of treatment groups from the researchers conducting the endpoint evaluations (and, in rare but ideal situations, from all research personnel and technicians). Concern regarding blinding may

Domain and Metric	*Criteria*
	be attenuated when outcome measures are more objective (e.g., as is the case of obtaining organ weights) or measurement is automated using computer-driven systems (e.g., as is the case in many behavioral assessments). For some outcomes, particularly histopathology assessment, outcome assessors are not blind to study group as they require comparison to the control to appropriately judge the outcome, but additional measures such as multiple levels of independent review by trained pathologists can minimize this potential bias. In animal studies, blinding of study group during the course of the study is often not possible for animal welfare considerations and the need to determine if treated animals are affected relative to controls in a treatment or dose-dependent manner (examples include clinical observations and histopathologic assessment of non-neoplastic lesions). Knowledge and tracking of higher exposed animals may also be part of animal welfare practices designed to avoid suffering associated with overtly toxic treatment doses. Under some conditions it is unlikely that blinding of research personnel during the course of a study can be fully achieved. However, animal studies are in general more tightly controlled than human studies and additional measures may be taken to reduce the risk of bias, such as the generation and use of standard operating procedures, training, and randomized husbandry or handling practices (e.g., placement in the animal room, necropsy order, etc.).
Confounding/Variable Control Control for independent variables across experimental groups	In a good study, outside of the chemical exposure of interest, all variables will be controlled for and consistent across experimental groups. Concern regarding additional variables, introduced intentionally or unintentionally, may be mitigated by knowledge or inferences regarding the likelihood and extent to which the variable can influence the endpoint(s) of interest. A very important example to consider is whether the exposure was sufficiently controlled to attribute the effects of exposure to the compound of interest alone. Generally, well-conducted exposures will not have any evidence of co-exposures and will include experimental controls that minimize the potential for confounding (e.g., use of a suitable vehicle control). Other examples of variables that may be uncontrolled or inconsistent across experimental groups include: protective or toxic factors that could mask or exacerbate effects; diet composition; surgical procedures (e.g., ovariectomy).
Reporting or Attrition Bias Lack of selective data reporting and unaccounted for loss of animals	In a good study, information is reported on all pre-specified outcomes and comparisons for all animals, across treatment groups and scheduled sacrifices. Aspects to consider include whether all study animals were accounted for in the results (if not, are explanations, such as death while on study, and adjustments provided), and whether expected comparisons or certain groups were excluded from the analyses. In

Domain and Metric	*Criteria*
	some studies, the outcomes evaluated must be inferred (e.g., a suite of standard measures in a guideline study). *Note*: This metric does not address whether quantitative data were reported, nor considers appropriateness of statistical analyses.
Exposure Methods Sensitivity Characterization of the exposure to the compound of interest	Consider whether there are notable issues that raise doubt about the accuracy of the exposure levels, or of exposure to the compound of interest. Depending on the chemical being assessed, this may include considering factors such as: the stability and composition (e.g., purity; isomeric composition) of the test article; exposure generation and analytic verification methods (including whether the tested levels and spacing between exposure groups is resolvable using current methods); and details of exposure methods (e.g., inhalation chamber type; gavage volume). In some cases, exposure biomarkers in blood, urine, or tissues of treated animals can mitigate concerns regarding inaccurate dosing (dependent on the validity of the biomarker for the chemical of interest). *Note*: While this may identify uncertainties in dose-response, such uncertainties are not considered a reason for exclusion from Hazard ID.
Utility of the exposure design for the endpoint of interest	Based on the known or presumed biological progression of the outcomes being evaluated, consider whether there are notable concerns regarding the timing, frequency, or duration of exposure. For example, better developmental studies will cover a greater proportion of the developmental window thought to be critical to the system of interest, while better studies for assessing cancer or other chronic outcomes will be of longer duration. Studies that expose animals infrequently or sporadically, or, conversely, on a continuous basis (which, depending on the exposure level, can impact food/ water consumption, sleep cycles, or pregnancy/ maternal care), might introduce additional complications.
Outcomes Measures and Results Display Sensitivity and specificity of the endpoint evaluations	Consider whether there are notable concerns about aspects of the procedures for, or the timing of, the endpoint evaluations. Based on the endpoint evaluation protocol used for the endpoints of interest, specific considerations will typically include: • Concerns regarding the sensitivity of the specific protocols for evaluating the endpoint of interest (i.e. assays can differ dramatically in terms of their ability to detect effects), and/or their timing (i.e. the age of animals at assessment can be critical to the appropriateness and sensitivity of the evaluation). This includes both overestimates or underestimates of the true effect, as well as a much higher (or lower) probability for detecting the effect(s) being assessed. • Concerns regarding the specificity and validity of the protocols. This includes the use of appropriate protocol controls to rule out non-specific effects, which can often be inferred from established guidelines or historical assay data. It may be

Domain and Metric	***Criteria***
	considered useful for insensitive, complex, or novel protocols to include positive and/or negative controls. • Concerns regarding adequate sampling. This includes both the experimental unit (e.g., litter; animal) and endpoint (e.g., number of slides evaluated in a histological analysis). This is typically inferred from historical knowledge of the assay or comparable assays. Although concerns about sample size may be noted, this factor is generally not a reason for a rating of critically deficient.
Usability and transparency of the presented data	Consider whether the results are analyzed or presented in a way that limits concerns regarding the credibility of the findings. Items that will typically be important to consider include: • Concern that the level of detail provided does not allow for an informed interpretation of the results (e.g., authors' conclusions without quantitative data; discussing neoplasms without distinguishing between benign and malignant tumors; not presenting variability). • Concern that the way in which the data were analyzed, compared, or presented is inappropriate or misleading. Examples include: failing to control for litter effects (e.g., when presenting pup data rather than the preferred litter data); pooling results from males and females or across lesion types; failing to address observed or presumed toxicity (e.g., in assessed animals; in dams) when exposure levels are known or expected to be highly toxic; incomplete presentation of the data (e.g., presenting continuous data as dichotomized); or non-preferred display of results (e.g., using a different readout than is expected for that assay). The evaluator should support how or why, and to what extent, this might mislead interpretations. *Notes:* Evaluating concerns regarding the statistical methods may require review by statistical experts. Missing information related to this metric should typically be requested from study authors.
Other (optional)	Example 1: Control for other threats to internal validity, e.g., animal husbandry concerns, reports of pre-dosing toxicity or infection, etc. Example 2: Concern for sensitivity of the animal model, e.g., demonstrated evidence of differences in model (e.g., species, sex, strain) sensitivity. This does not address the human relevance of the animal model, which is addressed elsewhere (PECO or during evidence synthesis)

5. DATA EXTRACTION OF STUDY METHODS AND RESULTS

Data extraction of animal study results was carried out using Health Assessment Workspace Collaborative (HAWC), a free and open source web-based software application[5], and data extraction of epidemiological results was carried out using Microsoft Word. For animal studies, the data extraction results will be available for download from HAWC in Excel format upon publication. Data extraction was performed by one member of the evaluation team and checked by 1-2 other members. Any discrepancies in data extraction were resolved by discussion or consultation. Once data were verified, they were "locked" to prevent accidental changes. Digital rulers were used to extract numerical information from figures, e.g., WebPlotDigitizer, [HYPERLINK "http://arohatgi.info/WebPlotDigitizer/"].

Missing data from individual animal and in vitro studies was generally not sought. However, more attempts were made for missing data, e.g., missing group size or variance descriptor (standard deviation/standard error from certain animal studies) on an otherwise well-reported and well-conducted animal study. Routine attempts were made to obtain missing information from epidemiologic studies, focusing on information required to conduct a meta-analysis. Outreach to study authors was documented and considered unsuccessful if researchers did not respond to an email or phone request within one month of the attempt to contact. Missing information to assess risk of bias and sensitivity for animal and epidemiologic studies was routinely sought as described above.

5.1. Standardizing administered dose levels

Dose levels provided in animal studies are presented as mg/kg-day. For studies in which dose levels are presented only as ppm (e.g. dietary exposure studies), dose conversions to mg/kg-day were made using US EPA default food or water consumption rates and body weights (using subchronic age and experiment duration) for the species/strain and sex of the animal of interest [ADDIN EN.CITE <EndNote><Cite><Author>EPA</Author><Year>1994</Year><RecNum>14</RecNum><DisplayText>(US EPA, 1994, 1988)</DisplayText><record><rec-number>14</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509461355">14</key></foreign-keys><ref-type name="Report">27</ref-type><contributors><authors><author>US EPA,</author></authors></contributors><titles><title>Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry</title></titles><pages>1-409</pages><dates><year>1994</year></dates><pub-location>Research Triangle Park,

[5] Health Assessment Workspace Collaborative (HAWC): A Modular Web-based Interface to Facilitate Development of Human Health Assessments of Chemicals. [HYPERLINK "https://hawcproject.org/portal/"].

NC</pub-location><publisher>U.S. Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office</publisher><isbn>EPA/600/8-90/066F</isbn><label>6488</label><work-type>EPA Report</work-type><urls><related-urls><url>https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=71993&CFID=51174829&CFTOKEN=25006317</url></related-urls></urls><language>English</language></record></Cite><Cite><Author>EPA</Author><Year>1988</Year><RecNum>15</RecNum><record><rec-number>15</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509461391">15</key></foreign-keys><ref-type name="Report">27</ref-type><contributors><authors><author>US EPA,</author></authors></contributors><titles><title>Recommendations for and documentation of biological values for use in risk assessment</title></titles><pages>1-395</pages><dates><year>1988</year></dates><pub-location>Cincinnati, OH</pub-location><publisher>U.S. Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment</publisher><isbn>EPA/600/6-87/008</isbn><label>64560</label><urls><related-urls><url>http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=34855</url></related-urls></urls><language>English</language></record></Cite></EndNote>]. The strain-specific and sex-specific values were used when available; for strains that are not available, the "other" values were used. Dose levels in mg/kg-day can vary for a given ppm across different studies if the studies use different species/strain or sex of animals, or both, that are assumed to have different food or water consumption rates. For studies in which dosing is through the feed, conversion was first made from food ppm to mg/kg-day. Unless otherwise reported by study authors, the background level in experimental animal studies was assumed 0 ppm (0 mg/kg-day).

6. SYNTHESIS WITHIN LINES OF EVIDENCE

For each health outcome, the human evidence and the animal evidence were synthesized separately for each chemical, augmenting each with informative subsets of mechanistic data. Each synthesis considered aspects of an association that may suggest causation: consistency, exposure–response relationship, strength of association, temporal relationship, biological plausibility, coherence, and "natural experiments" in humans [ADDIN EN.CITE <EndNote><Cite><Author>EPA</Author><Year>1994</Year><RecNum>45</RecNum><Suffix>`,'` , §2.1.3</Suffix><DisplayText>(US EPA, 2005,', §2.5; 1994,', §2.1.3)</DisplayText><record><rec-number>45</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509465929">45</key></foreign-keys><ref-type name="Report">27</ref-type><contributors><authors><author>US

EPA,</author></authors></contributors><titles><title>Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry</title></titles><pages>1-409</pages><dates><year>1994</year></dates><pub-location>Research Triangle Park, NC</pub-location><publisher>U.S. Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office</publisher><isbn>EPA/600/8-90/066F</isbn><label>6488</label><work-type>EPA Report</work-type><urls><related-urls><url>https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=71993&CFID=51174829&CFTOKEN=25006317</url></related-urls></urls><language>English</language></record></Cite><Cite><Author>EPA</Author><Year>2005</Year><RecNum>16</RecNum><Suffix>`,'`, §2.5</Suffix><record><rec-number>16</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509462499">16</key></foreign-keys><ref-type name="Report">27</ref-type><contributors><authors><author>US EPA,</author></authors></contributors><titles><title>Guidelines for carcinogen risk assessment</title></titles><pages>1-166</pages><dates><year>2005</year></dates><pub-location>Washington, DC</pub-location><publisher>U.S. Environmental Protection Agency, Risk Assessment Forum</publisher><isbn>EPA/630/P-03/001F</isbn><label>86237</label><work-type>EPA Report</work-type><urls><related-urls><url>http://www2.epa.gov/osa/guidelines-carcinogen-risk-assessment</url></related-urls></urls><language>English</language></record></Cite></EndNote>]. This leads to a distinction between *conflicting evidence* (unexplained positive and negative results in similarly exposed human populations or in similar animal models) and *differing results* (mixed results attributable to differences between human populations, animal models, or exposure conditions) [ADDIN EN.CITE <EndNote><Cite><Author>EPA</Author><Year>2005</Year><RecNum>16</RecNum><Suffix>`,'`, §2.5</Suffix><DisplayText>(US EPA, 2005,', §2.5)</DisplayText><record><rec-number>16</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509462499">16</key></foreign-keys><ref-type name="Report">27</ref-type><contributors><authors><author>US EPA,</author></authors></contributors><titles><title>Guidelines for carcinogen risk assessment</title></titles><pages>1-166</pages><dates><year>2005</year></dates><pub-location>Washington, DC</pub-location><publisher>U.S. Environmental Protection Agency, Risk Assessment Forum</publisher><isbn>EPA/630/P-03/001F</isbn><label>86237</label><work-type>EPA Report</work-type><urls><related-urls><url>http://www2.epa.gov/osa/guidelines-carcinogen-risk-assessment</url></related-urls></urls><language>English</language></record></Cite></EndNote>]. Each synthesis seeks

to reconcile ostensible inconsistencies between studies, taking into account differences in study methods and quality.

Biological significance of available evidence was considered more relevant to the assessment than statistical significance. Statistical significance (as reported by p-values, etc.) provides no evidence about effect size or biological significance, and lack of statistical significance will not be automatically interpreted as evidence of no effect. For both human and animal evidence, if the available data supports it, additional analyses across studies (such as meta-analysis) may also be conducted.

6.1. Syntheses of human and animal health effects evidence

To assess the weight of evidence regarding the potential for chemical exposure to cause a particular health effect, the syntheses of the human and animal health effects evidence focused on describing aspects of the evidence that best inform the casual interpretations. These syntheses were based primarily on studies of high and medium confidence. Low confidence studies were generally used to evaluate consistency, or if no or few higher confidence studies are available. If possible, the analysis included examination of results stratified by any or all of the following: study confidence classification (or specific issues within confidence evaluation domains), exposure level, sensitivity (e.g., low versus high), and other factors that may have been identified in the preliminary analysis plan (e.g., sex, lifestage, or other demographic). The number of studies and the differences encompassed by the studies determined the extent to which specific types of factors can be examined to stratify study results.

Syntheses clearly articulate the strengths and the weaknesses of the available evidence in the context of the considerations described in Table 10, which adapt the considerations first discussed by Austin Bradford Hill [ADDIN EN.CITE <EndNote><Cite><Author>Hill</Author><Year>1965</Year><RecNum>46</RecNum><DisplayText>(Hill, 1965)</DisplayText><record><rec-number>46</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509466265">46</key></foreign-keys><ref-type name="Journal Article">17</ref-type><contributors><authors><author>Hill, A. B.</author></authors></contributors><titles><title>The environment and disease: Association or causation?</title><secondary-title>Proceedings of the Royal Society of Medicine</secondary-title><alt-title>Proc R Soc Med</alt-title></titles><periodical><full-title>Proceedings of the Royal Society of Medicine</full-title><abbr-1>Proc R Soc Med</abbr-1></periodical><alt-periodical><full-title>Proceedings of the Royal Society of Medicine</full-title><abbr-1>Proc R Soc Med</abbr-1></alt-periodical><pages>295-300</pages><volume>58</volume><number>5</number><dates><year>1965</year></dates><isbn>ISSN 0035-9157</isbn><accession-num>14283879</accession-

num><label>71664</label><urls></urls><language>English</language></record></Cite></End Note>] and which are generally most informative for drawing conclusions on the integrated evidence. Table 10 also addresses consideration of the available mechanistic evidence. Because the human and animal syntheses typically provide a foundation for the evidence integration decisions, these syntheses include consideration of the ways in which mechanistic information might inform interpretations (e.g., address remaining uncertainties or data gaps) regarding the available human and animal health effect data. In addition, to the extent the data allow, the syntheses discuss factors relating to potential susceptible populations, based on demographics, genetic variability, lifestage, health status, behaviors or practices, social determinants, and other pollutant exposures.

Table [SEQ Table * ARABIC]. Primary considerations for synthesizing the human and, separately, animal health effect evidence

Consideration	Description
Consistency	Repeated findings across different studies or experiments increase the evidence strength. When inconsistencies exist, the evaluation considers study confidence and whether results were "conflicting" or "differing" [ADDIN EN.CITE <EndNote><Cite><Author>US EPA</Author><Year>2005</Year><RecNum>16</RecNum><DisplayText>(US EPA, 2005)</DisplayText><record><rec-number>16</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509462499">16</key></foreign-keys><ref-type name="Report">27</ref-type><contributors><authors><author>US EPA,</author></authors></contributors><titles><title>Guidelines for carcinogen risk assessment</title></titles><pages>1-166</pages><dates><year>2005</year></dates><pub-location>Washington, DC</pub-location><publisher>U.S. Environmental Protection Agency, Risk Assessment Forum</publisher><isbn>EPA/630/P-03/001F</isbn><label>86237</label><work-type>EPA Report</work-type><urls><related-urls><url>http://www2.epa.gov/osa/guidelines-carcinogen-risk-assessment</url></related-urls></urls><language>English</language></record></Cite></EndNote>]. Conflicting results decrease evidence strength. *Stronger human evidence*: evidence in different populations and study designs *Stronger animal evidence*: evidence in different species and strains, by different researchers
Biological gradient (dose-response[b])	Increases in risk, or in the frequency or severity of effects, with increasing exposure (e.g., concentration; duration) increase the evidence strength. This can reflect simple or complex (i.e., nonlinear) relationships. A lack of dose-response does not necessarily decrease evidence strength, but this may be appropriate after considering other studies and known biology.
Strength (effect magnitude) and precision	Given what is known about the health outcome, larger effect sizes or higher relative risks, particularly for rare or severe effects, are more concerning. Although small effect sizes are not grounds to dismiss an association, the

Consideration	Description
	evaluation of evidence strength may consider variability, historical data, or bias to assess the likelihood that effects are due to other explanations.
Biological plausibility	Supporting mechanistic evidence (e.g., associations with precursors or biomarkers related to effects; changes in established biological pathways or a theoretical mode-of-action) increases evidence strength. While a lack of mechanistic understanding does not decrease evidence strength, this can occur if findings demonstrate that effects are unlikely to occur. *Human evidence*: studies in exposed humans or appropriately exposed human cells *Animal evidence*: studies in exposed animals or appropriately exposed animal cells
Coherence[c]	Findings across the database that fit into a consistent pattern as a whole and hold together (e.g., similarity in results for related effects within an organ system, or across systems; a temporal or dose-dependent progression of linked effects of increasing severity) increase evidence strength. Conversely, an observed lack of changes that would be expected to occur (e.g., in parallel; subsequently) with the effect of interest could decrease evidence strength. Coherence is informed by the known biological development of the health effect in question, as well as toxicokinetic/dynamic understanding of the chemical or related chemicals [d].
Natural experiments	*Human evidence only*: Reductions in effect that occur after a clear reduction in exposure. Although rare, this can provide compelling, highly persuasive, evidence.
Temporality	*Human evidence only*: The exposure occurs before the effect (this issue is considered in the evaluation of exposure measures for each study)

[a]These ideas build upon the discussion for assessing causality of disease in [HYPERLINK \l "_ENREF_20" \o "Hill, 1965 #46"], although there are some differences in the use or interpretations of the terms.

[b] While humans are "exposed" and not "dosed", and nor are animals "dosed" via inhalation, "dose-response" is used for convention, although it is acknowledged that "exposure-response" may be more appropriate in many contexts.

[c] There is a clear overlap in the use of mechanistic evidence to interpret coherence (e.g., informing the relatedness or comparability of potentially coherent health findings) and biological plausibility. The available mechanistic information is synthesized separately and considered during the subsequent step of evidence integration (see Section 10).

[d] Although it is not separately listed, Hill's consideration of "analogy" (information for a similar but different association that supports causation) is indirectly encompassed by the evaluation of coherence during the review of environmental health studies; however, this use of analogous chemicals or exposure scenarios is less common.

For epidemiology evidence, the primary considerations used to inform causality and explore alternative explanations in the synthesis text were consistency (considering risk and direction of potential bias and sensitivity), biological gradient, strength (effect estimate magnitude and precision), coherence, natural experiments, and temporality. For experimental animal evidence, the primary considerations for the synthesis were consistency, dose-response gradient, strength (effect magnitude and precision), and coherence. Consistency often represented one of the most influential considerations, and the syntheses specifically emphasized observations across populations (e.g., location) or exposure scenarios in human studies, and across laboratories or populations (e.g., species, strain) in animal studies. When discussing the consistency of a set of study results, it is important to try to differentiate between conflicting evidence (unexplained positive and negative results in similarly exposed human populations or in similar animal models) and differing results (mixed results attributable to differences between human populations, animal models, or exposure conditions) [ADDIN EN.CITE <EndNote><Cite><Author>EPA</Author><Year>2005</Year><RecNum>16</RecNum><DisplayText>(US EPA, 2005)</DisplayText><record><rec-number>16</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509462499">16</key></foreign-keys><ref-type name="Report">27</ref-type><contributors><authors><author>US EPA,</author></authors></contributors><titles><title>Guidelines for carcinogen risk assessment</title></titles><pages>1-166</pages><dates><year>2005</year></dates><pub-location>Washington, DC</pub-location><publisher>U.S. Environmental Protection Agency, Risk Assessment Forum</publisher><isbn>EPA/630/P-03/001F</isbn><label>86237</label><work-type>EPA Report</work-type><urls><related-urls><url>http://www2.epa.gov/osa/guidelines-carcinogen-risk-assessment</url></related-urls></urls><language>English</language></record></Cite></EndNote>]. Some study results that appear to be inconsistent may be explained by potential biases or other attributes that affect the ability of a study to detect a true effect (study sensitivity) or that can artificially distort and create the appearance of an effect that does not truly exist, resulting in variations in the degree of confidence accorded to the study results. Additionally, the interpretation of the consistency of the evidence and the magnitude of the reported effects emphasized biological significance as more relevant to the assessment than statistical significance. Statistical significance (as reported by p-values) provides no evidence about effect size or biological significance, and lack of statistical significance was not automatically interpreted as evidence of no effect.

6.2. Considerations for pursuing a narrative or quantitative evidence synthesis

Heterogeneity within the available evidence was considered when determining whether to calculate an overall estimate of effect (meta-analysis) or other type of quantitative analysis. The principal characteristics evaluated for heterogeneity across eligible studies include the following:

Human Studies:

- Study design (e.g., cross-sectional, cohort)
- Details on how participants were classified into exposure groups, if any (e.g., quartiles of exposure concentration)
- Details on source of exposure data (e.g., questionnaire, area monitoring, biomonitoring)
- Exposure levels for each group
- Health outcome(s) reported
- Conditioning variables in the analysis or through population selection (e.g., variables considered confounders)
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, ability to access raw data
- Variation in degree and direction of risk of bias at individual study level

Animal Studies:

- Experimental design (randomized or not, acute or chronic, multigenerational, etc.)
- Animal model used (species, strain, sex, and genetic background)
- Age of animals (at start of treatment, mating, and/or pregnancy status)
- Developmental stage of animals at treatment and outcome assessment
- Dose levels, frequency of treatment, timing, duration, and exposure route
- Health outcome(s) reported
- Type of data (e.g., continuous or dichotomous), statistics presented in paper, ability to access raw data

6.3. Characterizing confidence in the overall body of human or animal evidence

A standardized set of descriptors was used to describe separately the overall confidence in the human or animal evidence for a given health outcome. Overall confidence determinations were

reached using a structured framework using similar considerations as the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach for evaluating certainty in the evidence [ADDIN EN.CITE ADDIN EN.CITE.DATA]. Tables 11 and 12 provide descriptions of the level of evidence needed to meet each classification for human evidence and animal evidence, respectively. Briefly, the terms **Robust** and **Moderate** describe evidence that supports a hazard. These terms are differentiated by the quantity and quality of information available to rule out alternative explanations for the results. **Slight** evidence includes situations in which there is some evidence that supports a hazard but a conclusion of **Moderate** does not apply. **Indeterminate** describes a situation where there are no studies available for that evidence stream or the evidence is inconsistent and cannot provide a basis for making a conclusion in either direction. **Compelling evidence of no effect** represents a situation where extensive evidence across a range of populations and exposures identified no association (uncommon).

Table [SEQ Table * ARABIC]. Framework for classification of evidence from studies in humans.

Extent of support for hazard	Within-stream strength of evidence conclusion	Description
Supports hazard	*Robust ... human evidence of an effect*	A set of *high* or *medium confidence* independent studies reporting an association between the exposure and the health outcome, with reasonable confidence that alternative explanations, including chance, bias, and confounding, are ruled out across studies. The set of studies is primarily consistent, and there are reasonable explanations when results differ; an exposure-response gradient is demonstrated; and the set of studies includes varied populations. Additional supporting evidence, such as associations with biologically-related endpoints in human studies (coherence) or large estimates of risk, may increase confidence but are not required. Selective reporting and publication bias are not a reasonable explanation for results. In exceptional circumstances, a finding in one study may be considered to be *robust*, even when other studies are not available (e.g., analogous to the finding of angiosarcoma, an exceedingly rare liver cancer, in the vinyl chloride industry). Mechanistic evidence from exposed humans or human cells, if available, may add support informing considerations such as exposure-response, temporality, coherence, and MOA, thus raising the level of certainty to *robust* for a set of studies that otherwise would be described as *moderate*.
	Moderate ... human evidence of an effect	A smaller number of studies (at least one *high* or *medium confidence* study with supporting evidence), or with some heterogeneous results, that do not reach the degree of confidence required for *robust*. There is some consistent evidence of an association, but alternative explanations, including chance, bias and confounding, have not been ruled out. Associations with related endpoints including mechanistic evidence from exposed humans or human cells, if available, may add support informing

Extent of support for hazard	Within-stream strength of evidence conclusion	Description
		considerations such as exposure-response, temporality, coherence, and MOA, thus raising the level of certainty to *moderate* for a set of studies that otherwise would be described as borderline *moderate/slight*.
Could support hazard or no hazard	*Slight ... human evidence of an effect*	One or more studies reporting an association between exposure and the health outcome, where alternative explanations exist. In general, only *low confidence* studies may be available, or considerable heterogeneity across studies may exist, and a MOA is not understood. Strong biological support from mechanistic studies in exposed humans or human cells may also be independently interpreted as *slight*. More rarely, a single *high confidence* study that is the initial evaluation of the study question (e.g., a hypothesis-generating vs. hypothesis-testing analysis) would also be described as *slight*. This category serves primarily to encourage additional study where evidence does exist that might provide some support for an association, but for which the evidence does not reach the degree of confidence required for *moderate*.
	Indeterminate ... human evidence of an effect	No studies available in humans or only a set of weak studies that are not consistent or are not informative to the hazard question under evaluation.
Supports no hazard	*Compelling evidence of no effect ... in human studies*	Several *high confidence* studies, showing consistently null results (for example, an odds ratio of 1.0) ruling out alternative explanations including chance, bias, and confounding with reasonable confidence. Each of the studies should have used an optimal outcome and exposure assessment and adequate sample size (specifically for higher exposure groups and for sensitive populations). The set as a whole should include the full range of levels of exposures that human beings are known to encounter, an evaluation of an exposure-response gradient, and at-risk populations and lifestages, and should be mutually consistent in not showing any indication of effect at any level of exposure.

Table [SEQ Table * ARABIC]. Framework for classification of evidence from studies in animals.

Extent of support for hazard	Within-stream strength of evidence conclusion	Description
Supports hazard	*Robust ... evidence of an effect in animals*	Consistent evidence for effects in animals has been observed in *high-to-medium-confidence* experiments[a] of varied design; any inconsistent evidence (evidence that cannot be reasonably explained by the respective study design or differences in animal model) is from a set of weaker experiments. The set of experiments supporting a hazard includes consistent findings of adverse or toxicologically significant effects across multiple laboratories or species, and the design of the experiments can reasonably rule out the potential for nonspecific effects (e.g., toxicity) to have resulted in the findings. Multiple lines of additional evidence in the set of experiments support a causal association, such as coherent effects across multiple related endpoints (may include mechanistic endpoints); an unusual magnitude of effect, rarity, age at onset, or severity; a strong dose-response relationship; and/or consistent observations across exposure scenarios (e.g., route; timing; duration), sexes, or animal strains. Mechanistic data in animals or animal cells that address the above considerations or that provide experimental support for a MOA that defines a causal relationship with reasonable confidence may raise the level of certainty to *robust* for evidence that otherwise would be described as *moderate* or, exceptionally, *slight/indeterminate*.
	Moderate ... evidence of an effect in animals	A set of evidence that does not reach the degree of certainty required for *robust*, but which includes at least one *high* or *medium confidence* experiment and supporting information. Although the results are largely consistent, notable uncertainties remain regarding the causal nature of the observed association. However, while inconsistent evidence and/or evidence indicating nonspecific effects may exist, it is not sufficient to reduce or discount the level of concern regarding the positive findings from the supportive experiments. Additionally, the set of supportive experiments provide evidence supporting a causal association, such as consistent effects across laboratories or species; coherent effects across multiple related endpoints (may include mechanistic endpoints); an unusual magnitude of effect, rarity, age at onset, or severity; a strong dose-response relationship; and/or consistent observations across exposure scenarios (e.g., route; timing; duration), sexes, or animal strains. Mechanistic data in animals or animal cells that address the above considerations or that provide information supporting an association between exposure and effect with reasonable confidence may raise the level of certainty to *moderate* for evidence that otherwise would be described as *slight*.
Could support hazard or no hazard	*Slight ... evidence in animals*	A set of experiments for which none of the other conclusions apply. This includes situations where only *low confidence* experiments are available and a MOA is not understood. Strong biological support from mechanistic studies in exposed animals or animal cells may also be independently interpreted as *slight*. Notably, to encourage additional research, it is important to describe situations where evidence does exist that might provide some support for an association, but for which the evidence is insufficient for a conclusion of *moderate*.

Extent of support for hazard	Within-stream strength of evidence conclusion	Description
	Indeterminate ...evidence in animals	No animal studies were available, or a set of *low confidence* animal studies exist that are not reasonably consistent or are not informative to the hazard question under evaluation.
Supports no hazard	*Compelling evidence of no effect ... in animals*	A set of *high confidence* experiments examining the full spectrum of related endpoints within a type of toxicity, with multiple species, and testing a reasonable range of exposure levels and adequate sample size in both sexes, with none showing any indication of effects. The data are compelling in that the experiments have examined the range of scenarios across which health effects in animals could be observed, and an alternative explanation (e.g., inadequately controlled features of the studies' experimental designs) for the observed lack of effects is not available. The experiments were designed to specifically test for effects of interest, including suitable exposure timing and duration, post-exposure latency, and endpoint evaluation procedures, and to address potentially susceptible populations and lifestages.

[a]"Experiment" is used here to refer to measurements in a single cohort of exposed animals (e.g., a study that included separate evaluations of rats and of mice, or separate cohorts exposed at different lifestages, would be considered as multiple experiments). Conversely, two papers or studies that report on the same cohort of exposed animals (e.g., examining different endpoints) would not be considered to be separate experiments. This language is used to reduce confusion regarding the use of the term "study."

6.4. Mechanistic Information

Mechanistic information includes any experimental measurement related to an endpoint or outcome that informs the biological or chemical events associated with toxic effects. This includes virtually all in vitro studies, and may also include mechanistic data from human and animal studies. Mechanistic information can be used to inform the synthesis and integration of the health effects evidence for both hazard identification and dose-response. An agent may operate through multiple mechanistic pathways, even if one hypothesis dominates the literature [ADDIN EN.CITE <EndNote><Cite><Author>EPA</Author><Year>2005</Year><RecNum>16</RecNum><Suffix>`,'` , §2.4.3.3</Suffix><DisplayText>(US EPA, 2005,', §2.4.3.3)</DisplayText><record><rec-number>16</rec-number><foreign-keys><key app="EN" db-id="vpzara2f69w5wjesvxkxzf90efs2ztdrxdps" timestamp="1509462499">16</key></foreign-keys><ref-type name="Report">27</ref-type><contributors><authors><author>US EPA,</author></authors></contributors><titles><title>Guidelines for carcinogen risk assessment</title></titles><pages>1-166</pages><dates><year>2005</year></dates><pub-location>Washington, DC</pub-location><publisher>U.S. Environmental Protection Agency, Risk Assessment Forum</publisher><isbn>EPA/630/P-03/001F</isbn><label>86237</label><work-type>EPA Report</work-type><urls><related-urls><url>http://www2.epa.gov/osa/guidelines-

carcinogen-risk-assessment</url></related-urls></urls><language>English</language></record></Cite></EndNote>].

The synthesis of mechanistic data focuses on evidence most likely to be useful for augmenting the human or animal evidence with information on precursor events, evaluating the human relevance of animal results, or identifying susceptible populations and lifestages. Some examples of how mechanistic decisions can apply to overall hazard and evidence conclusions can be found in Table 13. Given the large number of diverse studies and the need for focused analyses informative to particular hazard questions, the mechanistic synthesis for a given health outcome generally required inclusion of only a subset of the most relevant mechanistic studies.

As the potential influence of the information provided by these studies varied depending on the hazard question(s), the rigor of the analyses likewise varied from cursory insights drawn from sets of unanalyzed results to detailed evaluations of a subset of the relevant, individual mechanistic studies. For example, there are a relatively large number of mechanistic studies pertaining to the male reproductive toxicity of phthalates, which allowed for the application of these data to evaluate human relevance of animal studies using an adverse outcome pathway framework. Comparatively, other health outcomes (i.e. female reproductive, developmental) had limited mechanistic data available, and therefore only a cursory overview was performed.

Mechanistic studies were identified in the comprehensive literature search for the chemical and preliminarily binned as "Studies with Supporting Data" during title/abstract screening. An iterative approach was used to determine which *in vitro* and other types of mechanistic studies were most important to summarize, based on factors such as directness or relevance of the model systems, concentrations tested, and robustness of the evidence in humans and animals.

Table [SEQ Table * ARABIC]. Examples of the interpretation and application of mechanistic evidence

Mechanistic inferences considered	Potential applications of the mechanistic synthesis
Biological plausibility: as applied herein, this describes information that either strengthens or weakens an interpretation of the likelihood of an association between exposure and the health effect. Thus, in some instances, differing levels of biological plausibility (or certainty) might be drawn. It is important to note that the lack of mechanistic data explaining an association is not used to discount observations from human or animal studies. The interpretation of biological plausibility considers the existing knowledge for how the health effect develops and can involve analyses of information at different levels of	***Evidence integration (within stream)*** • Observations of important mechanistic changes in exposed humans or animals that are plausibly associated with the health outcome in question can strengthen the confidence in the within stream health effect findings, particularly when the changes are observed in the same exposed population presenting the health effect. • The absence of expected mechanistic changes in an exposed population might diminish the plausibility of an association. This considers the sensitivity of the mechanistic changes and the potential contribution of alternate or unidentified mechanisms of toxicity. • Inconsistent evidence (i.e. heterogeneous results) across different animal species or human populations might be explainable by evidence that different mechanisms are operant in the different populations (e.g., evidence

Mechanistic inferences considered	Potential applications of the mechanistic synthesis
biological organization (e.g., molecular; tissue).	demonstrating that certain populations cannot metabolize a reactive metabolite; evidence that variability in gene expression correlates with variability in response).
Human relevance of findings in animals: in the absence of sufficient MOA information, effects in animal models are assumed to be relevant to humans, e.g. [HYPERLINK \l "_ENREF_47" \o "US EPA, 2005 #16"] (section 4.1.2.7) and [HYPERLINK \l "_ENREF_48" \o "US EPA, 2006 #50"]. For potential human health hazards supported by strong evidence from animal models, mechanistic evidence is considered in light of human relevance.	***Evidence integration (across stream)*** • If evidence establishes that the mechanisms underlying the animal response does not operate in humans, or that animal models do not suitably inform a specific human health outcome, this can support the view that the animal response is not relevant to humans. In these cases, the animal response provides neither an argument for nor an argument against an overall hazard judgment. • Observations of mechanistic changes in exposed humans that are similar or coherent with mechanistic or toxicological changes in experimental animals (and that are interpreted to be associated with the health outcome under evaluation) strengthen the human relevance of the animal findings.
Potential vulnerabilities: mechanistic understanding of how a health outcome develops, even without a full MOA, can clarify characteristics of important events (e.g., their presence or sensitivity across lifestages or across genetic variations) and help to identify vulnerable population groups. E.g. [HYPERLINK \l "_ENREF_47" \o "US EPA, 2005 #16"] (section 4.1.2.7)	***Susceptibility and dose-response analysis*** • Identification of life stages or groups likely at greatest risk can clarify hazard descriptions, including whether the most vulnerable populations have been adequately tested. • Knowledge of expected vulnerabilities can inform selection of studies for quantitative analysis, e.g., prioritizing studies including such populations. • When there is evidence of susceptibilities, but specific studies cannot be prioritized for quantitative analysis, susceptibility data may support refined human variability uncertainty factors or probabilistic uncertainty analyses.
Biological understanding, including the identification of precursor events: mechanistic data that reasonably describe how effects develop may clarify the exposure conditions expected to result in these effects. Further, well-studied MOAs can identify mechanistic precursor events linked qualitatively or quantitatively to apical health effect(s), increasing the strength of the hazard descriptor.	***Dose-response analysis*** MOA inferences can support the use of: • Particular dose-response models—e.g., biologically-based models, models integrating data across several related outcomes. • Proximal measures of exposure—e.g., external vs. internal metrics. • Improved characterization of responses—e.g., use of well-established precursors in lieu of direct observation of apical endpoints; combination of related outcomes (such as benign and malignant tumors resulting from the same MOA).

ABOUT THIS PROTOCOL

Contributors

Name	Affiliation
Xabier Arzuaga, Ph.D.	US EPA Office of Research and Development (ORD), National Center for Environmental Assessment (NCEA), Integrated Risk Information System (IRIS)
Brandiese E. J. Beverly, Ph.D. ‡	US EPA ORD/NCEA/IRIS
Todd Blessinger, Ph.D.	US EPA ORD/NCEA/IRIS
Christine Cai, M.S.	US EPA ORD/NCEA/IRIS
Glinda Cooper, Ph.D. †	US EPA ORD/NCEA/IRIS
Laura Dishaw, Ph.D.	US EPA ORD/NCEA/IRIS
Susan Y. Euling, Ph.D. ⊥	US EPA ORD/NCEA/IRIS
Audrey Galizia, Ph.D.	US EPA ORD/NCEA/IRIS
Andrew Hotchkiss, Ph.D.	US EPA ORD/NCEA/IRIS
Nagalakshmi Keshava, Ph.D.	US EPA ORD/NCEA/IRIS
Susan L. Makris, Ph.D.	US EPA ORD/NCEA/IRIS
Anuradha Mudipalli, Ph.D.	US EPA ORD/NCEA/IRIS'
Elizabeth Radke-Farabaugh, Ph.D.	US EPA ORD/NCEA/IRIS
Kristina A. Thayer, Ph.D.	US EPA ORD/NCEA/IRIS
Lily Wang, Ph.D.	US EPA ORD/NCEA/IRIS
Teneille D. Walker, Ph.D. §	US EPA ORD/NCEA/IRIS
James A. Weaver, Ph.D.	US EPA ORD/NCEA/IRIS
Erin E. Yost, Ph.D.	US EPA ORD/NCEA/IRIS
Contract support: Assisted in protocol development, literature screening, data extraction, risk of bias assessment and document production	
Louise Assem	ICF International, Inc.
Melissa Branigan	US EPA Student Services Contractor
Joseph Braun	Brown University
Heather Carlson-Lynch	SRC, Inc.
Anna Chen	US EPA Student Services Contractor
Ali Goldstone, MPH	ICF International, Inc.
Andrew Greenhalgh	US EPA Student Services Contractor
Pamela Hartman	ICF International, Inc.

Name	Affiliation
Evangela Matthews	US EPA Student Services Contractor
John Meeker	University of Michigan
Julie Melia	SRC, Inc.
Kim Osborn	ICF International, Inc.
Megan Ricardi	SRC, Inc.
Wendie Robbins	University of California, Los Angeles
Sarah Rosenberg	SRC, Inc.
Kelly Salinas	SRC, Inc.
Kimberly Zaccaria	SRC, Inc.

Current affiliations:

‡ US National Institute for Environmental Health Sciences, National Toxicology Program, Office of Health Assessment and Translation

† Innocence Project (https://www.innocenceproject.org/)

⊥ US EPA Office of Children's Health Protection

§ US Drug Enforcement Agency

REFERENCES

[ADDIN EN.REFLIST]

APPENDIX. Electronic Database Search Strategies

Appendix Table [SEQ Appendix_Table * ARABIC]. Database search strategy for phthalate epidemiological studies.

Database	Query strings
PubMed	(phthalate OR phthalates OR phthalic acid) AND (human OR case-control OR pregnancy OR cohort OR workers OR children OR survey)
Web of Science	(TS = "phthalic acid" OR TS = "phthalate" OR TS = "phthalates") AND (TS = "humans" OR TS = "human" OR TS = "case-control" OR TS = "pregnancy" OR TS = "cohort" OR TS = "workers" OR TS = "child" OR TS = "children" OR TS = "survey")
ToxLine	(phthalate OR phthalates OR phthalic acid) AND (human OR case-control OR pregnancy OR cohort OR workers OR children OR survey)

Search dates: 01/2017, 06/2016, 12/2015, 03/2015, 12/2014, 06/2014, 12/2013, 06/2013

Appendix Table [SEQ Appendix_Table * ARABIC]. Database search strategy for DIBP

Database (search date)	Keywords[a]
PubMed 07/2017 01/2017 06/2016 12/2015 06/2015 03/2014 02/2013	dibp OR (mibp AND phthalate) OR "diisobutylphthalate" OR "di-isobutyl phthalate" OR "84-69-5" OR "diisobutyl phthalate" OR "di(i-butyl)phthalate" OR "di-iso-butyl phthalate" OR "isobutyl phthalate" OR "phthalic acid diisobutyl ester" OR ("diisobutyl ester" AND phthalate) OR "1,2-benzenedicarboxylic acid bis(2-methylpropyl) ester" OR "1,2-benzenedicarboxylic acid 1,2-bis(2-methylpropyl) ester" OR "monoisobutyl phthalate" OR "mono(i-butyl)phthalate" OR "mono-iso-butyl phthalate" OR "phthalic acid monoisobutyl ester" OR "1,2-benzenedicarboxylic acid, mono(2-methylpropyl) ester" OR "2-[(2-methylpropoxy)carbonyl]benzoic acid" OR "1,2-benzenedicarboxylic acid, mono(2-methylpropyl) ester (9CI)" OR "isobutyl hydrogen phthalate" OR "1,2-benzenedicarboxylic acid 1-(2-methylpropyl) ester"
Web of Science 07/2017 01/2017 06/2016 12/2015 06/2015 03/2014 02/2013	TS=dibp OR (TS=mibp AND TS=phthalate) OR TS="diisobutylphthalate" OR TS="di-isobutyl phthalate" OR TS="84-69-5" OR TS="diisobutyl phthalate" OR TS="di(i-butyl)phthalate" OR TS="di-iso-butyl phthalate" OR TS="isobutyl phthalate" OR TS="phthalic acid diisobutyl ester" OR (TS="diisobutyl ester" AND TS=phthalate) OR TS="1,2-benzenedicarboxylic acid bis(2-methylpropyl) ester" OR TS="1,2-benzenedicarboxylic acid 1,2-bis(2-methylpropyl) ester" OR TS="monoisobutyl phthalate" OR TS="mono(i-butyl)phthalate" OR TS="mono-iso-butyl phthalate" OR TS="phthalic acid monoisobutyl ester" OR TS="1,2-benzenedicarboxylic acid, mono(2-methylpropyl) ester" OR TS="2-[(2-methylpropoxy)carbonyl]benzoic acid" OR TS="1,2-benzenedicarboxylic acid, mono(2-methylpropyl) ester (9CI)" OR TS="isobutyl hydrogen phthalate" OR TS="1,2-benzenedicarboxylic acid 1-(2-methylpropyl) ester"
Toxline 07/2017 01/2017 06/2016 12/2015 06/2015 03/2014	Split into 4 separate search strings: @TERM+@rn+84-69-5 @AND+mibp+phthalate @AND+"diisobutyl ester"+phthalate

02/2013	@OR+(dibp+"diisobutylphthalate"+"di-isobutyl+phthalate"+"diisobutyl+phthalate"+"di(i-butyl)phthalate"+"di-iso-butyl+phthalate"+"isobutyl+phthalate"+"phthalic+acid+diisobutyl+ester"+"1,2-benzenedicarboxylic+acid+bis(2-methylpropyl)+ester"+"1,2-benzenedicarboxylic+acid+1,2-bis(2-methylpropyl)+ester"+"monoisobutyl+phthalate"+"mono(i-butyl)phthalate"+"mono-iso-butyl+phthalate"+"phthalic+acid+monoisobutyl+ester"+"1,2-benzenedicarboxylic+acid,+mono(2-methylpropyl)+ester"+"2-[(2-methylpropoxy)carbonyl]benzoic+acid"+"1,2-benzenedicarboxylic+acid,+mono(2-methylpropyl)+ester+(9CI)"+"isobutyl+hydrogen+phthalate"+"1,2-benzenedicarboxylic+acid+1-(2-methylpropyl)+ester")
TSCATS2 07/2017 01/2017 06/2016 12/2015 06/2015 03/2014	(2000-) 84-69-5

Appendix Table [SEQ Appendix_Table * ARABIC]. Processes used to augment the search of core databases for DIBP

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
Manual search of citations from regulatory documents	CPSC. (2010). Toxicity Review for Diisobutyl phthalate (DIBP). Bethesda, MD: Consumer Product Safety Commission.	3/2014	9 citations added
	Australian Government (2017). Human health tier II assessment for C4-6 side chain transitional phthalates.	3/2017	3 citations added
	BAuA (2014). CLH report. Proposal for Harmonised Classification and Labelling. Diisobutylphthalate (DIBP).	3/2017	0 citations added
	European Chemicals Agency (2009) agreement of the member state committee on identification of diisobutyl phthalate (dibp) as a substance of very high concern.	3/2017	0 citations added
	European Chemicals Agency (2009) Member state committee support document for identification of diisobutyl phthalate as a substance of very high concern because of its CMR properties.	3/2017	1 citation added
	European Chemicals Agency (2014) Committee for Risk Assessment RAC Annex 1 Background document to the Opinion proposing harmonised classification and labelling at Community level of diisobutyl phthalate (DIBP).	3/2017	0 citations added
Web of Science forward search	Hannas BR, Lambright CS, Furr J, Howdeshell KL, Wilson VS, Gray LE Jr. (2011). Dose-response assessment of fetal testosterone production and gene expression levels	3/2014 3/2017	2 citations added 10 citations added

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	in rat testes following in utero exposure to diethylhexyl phthalate, diisobutyl phthalate, diisoheptyl phthalate, and diisononyl phthalate. Toxicol Sci. 123(1):206-16.		
	Saillenfait AM, Sabaté JP, Gallissot F. (2008). Diisobutyl phthalate impairs the androgen-dependent reproductive development of the male rat. Reprod Toxicol. 26(2):107-15.	3/2014 3/2017	1 citation added 6 citations added
	Ray B, D'Souza AS, Kumar V, Pugazhandhi B, D'Souza MR, Nayak D, Sushma RK, Shetty P, Singh H, Krishna L, Bhat KM, Rao AC, Chakraborti S, Kumar N, Saxena A. (2012). Ovarian development in Wistar rat treated prenatally with single dose diisobutyl phthalate. Bratisl Lek Listy. 113(10):577-82.	3/2014	0 citations added
	Kleinsasser NH, Wallner BC, Kastenbauer ER, Weissacher H, Harréus UA. (2001). Genotoxicity of di-butyl-phthalate and di-iso-butyl-phthalate in human lymphocytes and mucosal cells. Teratog Carcinog Mutagen. 21(3):189-96.	3/2014	1 citation added
	Sohn J; Kim S; Koschorreck J; Kho Y; Choi K. (2016). Alteration of sex hormone levels and steroidogenic pathway by several low molecular weight phthalates and their metabolites in male zebrafish (Danio rerio) and/or human adrenal cell (H295R) line. Journal of Hazardous Materials 320:45-54.	3/2017	0 citations added
	Borch J; Axelstad M; Vinggaard AM; Dalgaard M. (2006). Diisobutyl phthalate has comparable anti-androgenic effects to di-n-butyl phthalate in fetal rat testis. Toxicology Letters 163(3):183-190.	3/2017	9 citations added
Backward search Web of Science or manual	Hannas BR, Lambright CS, Furr J, Howdeshell KL, Wilson VS, Gray LE Jr. (2011). Dose-response assessment of fetal testosterone production and gene expression levels in rat testes following in utero exposure to diethylhexyl phthalate, diisobutyl phthalate, diisoheptyl phthalate, and diisononyl phthalate. Toxicol Sci. 123(1):206-16.	3/2014	1 citation added
	Saillenfait AM, Sabaté JP, Gallissot F. (2008). Diisobutyl phthalate impairs the androgen-dependent reproductive development of the male rat. Reprod Toxicol. 26(2):107-15.	3/2014	1 citation added
	Ray B, D'Souza AS, Kumar V, Pugazhandhi B, D'Souza MR, Nayak D, Sushma RK, Shetty P, Singh H, Krishna L, Bhat KM, Rao AC, Chakraborti S, Kumar N, Saxena A. (2012). Ovarian development in Wistar rat treated prenatally with single dose diisobutyl phthalate. Bratisl Lek Listy. 113(10):577-82.	3/2014	4 citations added

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	Kleinsasser NH, Wallner BC, Kastenbauer ER, Weissacher H, Harréus UA. (2001). Genotoxicity of di-butyl-phthalate and di-iso-butyl-phthalate in human lymphocytes and mucosal cells. Teratog Carcinog Mutagen. 21(3):189-96.	3/2014	2 citations added
	Sohn J; Kim S; Koschorreck J; Kho Y; Choi K. (2016). Alteration of sex hormone levels and steroidogenic pathway by several low molecular weight phthalates and their metabolites in male zebrafish (Danio rerio) and/or human adrenal cell (H295R) line. Journal of Hazardous Materials 320:45-54.	3/2017	1 citation added
	Wang X; Sheng N; Cui R; Zhang H; Wang J; Dai, J. (2017). Gestational and lactational exposure to di-isobutyl phthalate via diet in maternal mice decreases testosterone levels in male offspring. Chemosphere 172:260-267.	3/2017	3 citations added
	Borch J; Axelstad M; Vinggaard AM; Dalgaard M. (2006). Diisobutyl phthalate has comparable anti-androgenic effects to di-n-butyl phthalate in fetal rat testis. Toxicology Letters 163(3):183-190.	3/2017	0 citations added
References obtained during the assessment process	DIBP references in previous assessment or previously added to the HERO project page	4/2014	348 citations added
Search of online chemical assessment-related websites	Searched a combination of CASRNs and synonyms on the following databases: ACGIH (Hardcopy TLV booklet and documentation) ATSDR ([HYPERLINK "http://www.atsdr.cdc.gov/substances/index.asp"]) CalEPA Office of Environmental Health Hazard Assessment ([HYPERLINK "http://www.oehha.ca.gov/risk.html"]) OEHHA Toxicity Criteria Database ([HYPERLINK "http://www.oehha.ca.gov/tcdb/index.asp"]) Biomonitoring California-Priority Chemicals ([HYPERLINK "http://www.biomonitoring.ca.gov/chemicals/priority-chemicals)a"] Biomonitoring California-Designated Chemicals ([HYPERLINK "http://www.biomonitoring.ca.gov/chemicals/designated-chemicals)a"] Cal/Ecotox database ([HYPERLINK "http://www.oehha.ca.gov/scripts/cal_ecotox/CHEMLIST.ASP"])[b]	2/2013, 3/2014, 2/2017	17 citations added 8 citations added

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	OEHHA Fact Sheets ([HYPERLINK "http://www.oehha.ca.gov/public_info/facts/index.html"]) Non-cancer health effects Table (RELs) and Cancer Potency Factors (Appendix A and AppendixB) ([HYPERLINK "http://www.oehha.ca.gov/air/hot_spots/index.html"]) CPSC ([HYPERLINK "http://www.cpsc.gov"]) eChemPortal ([HYPERLINK "http://www.echemportal.org/echemportal/participant/page.action?pageID=9"]) Environment Canada – Search entire site if not found below: ([HYPERLINK "http://www.ec.gc.ca/default.asp?lang=En&n=ECD35C36"]) Toxic Substances Managed under CEPA ([HYPERLINK "http://www.ec.gc.ca/toxiques-toxics/Default.asp?lang=En&n=98E80CC6-1"]) Final Assessments ([HYPERLINK "http://www.ec.gc.ca/lcpe-cepa/default.asp?lang=En&xml=09F567A7-B1EE-1FEE-73DB-8AE6C1EB7658"]) Draft Assessments ([HYPERLINK "http://www.ec.gc.ca/lcpe-cepa/default.asp?lang=En&xml=6892C255-5597-C162-95FC-4B905320F8C9"]) EPA Acute Exposure Guideline Levels ([HYPERLINK "http://www.epa.gov/oppt/aegl/pubs/chemlist.htm"]) EPA – IRISTrack/New Assessments and Reviews [HYPERLINK "http://cfpub.epa.gov/ncea/iris/search/"] EPA NSCEP ([HYPERLINK "http://www.epa.gov/ncepihom/"]) EPA RfD/RfC and CRAVE meeting notes[c] EPA Science Inventory ([HYPERLINK "http://cfpub.epa.gov/si/"]) FDA ([HYPERLINK "http://www.fda.gov/"]) Federal Docket ([HYPERLINK "file:///C:/Users/riccardi/AppData/Local/Microsoft/AppData/Local/IRIS%20Tox%20Reviews/RDX/SearchHistory/LSP_201X/FOR%20INTERNAL%20USE%20ONLY%20-%20Search%20Table/www.regulations.gov"])		

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	Health Canada First Priority List Assessments ([HYPERLINK "http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl1-lsp1/index-eng.php"]) Health Canada Second Priority List Assessments ([HYPERLINK "http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl2-lsp2/index-eng.php"]) IARC ([HYPERLINK "http://monographs.iarc.fr/ENG/Classification/index.php"])[a] ITER (TERA database) ([HYPERLINK "http://www.tera.org/iter/"]) NAP – Search Site ([HYPERLINK "http://www.nap.edu/"]) NCI ([HYPERLINK "http://www.cancer.gov"]) National Institute for Environmental Health Sciences (NIEHS) [HYPERLINK "http://www.niehs.nih.gov/"] NICNAS (PEC only covered by eChemPortal) (http://www.nicnas.gov.au/chemical-information) NIOSH ([HYPERLINK "http://www.cdc.gov/niosh/topics/"]) NIOSHTIC 2 ([HYPERLINK "http://www2a.cdc.gov/nioshtic-2/"]) NTP - RoC, status, results, and management reports ([HYPERLINK "https://seek.niehs.nih.gov/texis/search/?mode=&opts=&pr=internet-all&dropXSL=html&sq=&prox=page&rorder=750&rprox=750&rdfreq=0&rwfreq=0&rlead=1000&rdepth=31&sufs=1&order=r&query=&mu=National+Toxicology+Program)a"] OSHA ([HYPERLINK "http://www.osha.gov/dts/chemicalsampling/toc/toc_chemsamp.html"]) RTECS [HYPERLINK "http://www.ccohs.ca/search.html"]		
	AIHA WEELs ([HYPERLINK "http://www.tera.org/OARS/WEEL.html"]) ERPGs ([HYPERLINK "https://www.aiha.org/get-involved/AIHAGuidelineFoundation/EmergencyResponsePlanningGuidelines/Documents/2014%20ERPG%20Values.pdf"])	2/2017	

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	CalEPA Drinking Water Notification Levels ([HYPERLINK "http://www.swrcb.ca.gov/drinking_water/certlic/drinkingwater/NotificationLevels.shtml"]) CHRIP ([HYPERLINK "http://www.safe.nite.go.jp/english/db.html"]) ECETOC publications ([HYPERLINK "http://www.ecetoc.org/publications"]) ECHA General site [HYPERLINK "http://echa.europa.eu/information-on-chemicals"] info on Registered Substances ([HYPERLINK "http://echa.europa.eu/information-on-chemicals/registered-substances"]) Information from the Existing Substances Regulation (ESR) including Final Risk Assessments ([HYPERLINK "http://echa.europa.eu/information-on-chemicals/information-from-existing-substances-regulation"]) Opinions of the Committee for Risk Assessment on proposals for harmonised classification and labelling [HYPERLINK "https://echa.europa.eu/opinions-of-the-committee-for-risk-assessment-on-proposals-for-harmonised-classification-and-labelling"] Opinions of the RAC adopted under specific ECHA's Executive Director requests		

	[HYPERLINK "https://echa.europa.eu/about-us/who-we-are/committee-for-risk-assessment/opinions-of-the-rac-adopted-under-specific-echa-s-executive-director-requests"] PACT – RMOA and hazard assessment activities [HYPERLINK "https://echa.europa.eu/addressing-chemicals-of-concern/substances-of-potential-concern/pact?p_p_id=viewsubstances_WAR_echarevsubstanceportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_pos=1&p_p_col_count=2&_viewsubstances_WAR_echarevsubstanceportlet_cur=1&_viewsubstances_WAR_echarevsubstanceportlet_delta=50&_viewsubstances_WAR_echarevsubstanceportlet_keywords=&_viewsubstances_WAR_echarevsubstanceportlet_advancedSearch=false&_viewsubstances_WAR_echarevsubstanceportlet_andOperator=true&_viewsubstances_WAR_echarevsubstanceportlet_orderByCol=synonymDynamicField_504&_viewsubstances_WAR_echarevsubstanceportlet_orderByType=asc"] Substance evaluation – CoRAP [HYPERLINK "https://echa.europa.eu/information-on-chemicals/evaluation/community-rolling-action-plan/corap-table"] European Union Risk Assessment Reports [HYPERLINK "https://ec.europa.eu/jrc/en/publications-list"] Health Canada – Search entire site [HYPERLINK "http://www.hc-sc.gc.ca/index-eng.php"] Health Canada Drinking Water Documents [HYPERLINK "http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/index-eng.php" \l "tech_doc"] Health Canada Indoor Air Quality Guidelines [HYPERLINK "http://healthycanadians.gc.ca/healthy-living-vie-saine/environment-environnement/air/guidelines-lignes-directrices-eng.php"] EPA CDAT [HYPERLINK "http://java.epa.gov/oppt_chemical_search/"] OPP [HYPERLINK "http://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1"]		

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	Japan Existing Chemical Data Base (JECDB) [HYPERLINK "http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp"] NTP 14th Report On Carcinogens: ([HYPERLINK "http://ntp.niehs.nih.gov/?objectid=03C9AF75-E1BF-FF40-DBA9EC0928DF8B15"]) OECD HPV/SIDS/IUCLID (cross-check with eChem) [HYPERLINK "http://webnet.oecd.org/hpv/ui/Search.aspx"] WHO Air quality guidelines ([HYPERLINK "http://www.who.int/phe/health_topics/outdoorair/outdoorair_aqg/en/"]) Indoor air quality guidelines ([HYPERLINK "http://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/2010/who-guidelines-for-indoor-air-quality-selected-pollutants"]) Drinking water quality guidelines ([HYPERLINK "http://www.who.int/water_sanitation_health/publications/2011/dwq_guidelines/en/"])		

[a]Website updated
[b]Current list unavailable online
[c]Not searched 2/2017 since it's not updated.

Appendix Table [SEQ Appendix_Table * ARABIC]. Database search strategy for DEP

Note: In 2017, the PubMed, Web of Science, and Toxline search strings for DEP were restructured to reduce the overall number of search strings, remove redundant search terms, and to search more broadly by (1) including plural forms of substance names, (2) expanding the list of terms in the Web of Science and Toxline search strings, and (3) searching all of PubMed for the substance names and synonyms, not just the non-indexed subset of PubMed.

Database (date searched)	Terms
Pubmed 07/2017 01/2017	((((("diethyl o-phthalate"[tw] OR "diethyl o-phthalates"[tw] OR "diethyl phthalate"[tw] OR "diethyl phthalates"[tw] OR "ethyl phthalate"[tw] OR "ethyl phthalates"[tw]) OR (DEP[tw] AND ("phthalate"[all] OR "phthalates"[all] OR "phthalate's"[all] OR "phthalated"[all] OR "phthalaten"[all] OR "phthalic acid"[all] OR "phthalic acids"[all]))) OR "84-66-2"[EC/RN Number]) AND ("2016/06/01"[Date - Publication] : "2017/01/31"[Date - Publication]))

Database (date searched)	Terms
PubMed 06/2016 01/2016 06/2015 02/2015 07/2014 8/31/13 10/2012	((("Diethyl o-phthalate"[tw] OR "Diethyl phthalate"[tw] OR "Ethyl phthalate"[tw]) OR (DEP[tw] AND (phthalate[All Fields] OR phthalate/1[All Fields] OR phthalate/2[All Fields] OR phthalate/25[All Fields] OR phthalate/adipate[All Fields] OR phthalate/aged[All Fields] OR phthalate/cellulose[All Fields] OR phthalate/dialkoxyalkyl[All Fields] OR phthalate/ethanol[All Fields] OR phthalate/ferrocene[All Fields] OR phthalate/goethite[All Fields] OR phthalate/kg[All Fields] OR phthalate/mg[All Fields] OR phthalate/ml[All Fields] OR phthalate/naoh[All Fields] OR phthalate/toxicity[All Fields] OR phthalate/water[All Fields] OR phthalate's[All Fields] OR phthalated[All Fields] OR phthalaten[All Fields] OR phthalates[All Fields] OR phthalates/kg/day[All Fields] OR phthalates/toxicity[All Fields] OR phthalates'[All Fields]))) NOT medline[sb]) OR "84-66-2"[EC/RN Number]

Web of Science 07/2017 01/2017	((TS="DEP" AND TS="phthalat*") OR TS="1,2-benzenedicarboxylic acid, diethyl ester" OR TS="diethyl 1,2-benzenedicarboxylate" OR TS="diethyl o-phthalate" OR TS="diethyl o-phthalates" OR TS="diethyl phthalate" OR TS="diethyl phthalates" OR TS="di-n-ethyl phthalate" OR TS="di-n-ethyl phthalates" OR TS="ethyl phthalate" OR TS="ethyl phthalates" OR TS="phthalic acid, diethyl ester" OR TS="unimoll da" OR TS="solvanol" OR TS="placidol e" OR TS="phthalol" OR TS="palatinol a" OR TS="neantine" OR TS="anozol") AND (TS="chronic" OR TS="immun*" OR TS="lymph*" OR TS="neurotox*" OR TS="toxicokin*" OR TS="pharmacokin*" OR TS="biomarker*" OR TS="neurolog*" OR TS="subchronic" OR TS="pbpk" OR TS="epidemiolog*" OR TS="acute" OR TS="subacute" OR TS="ld50" OR TS="lc50" OR TS="inhal*" OR TS="pulmon*" OR TS="nasal" OR TS="lung*" OR TS="respir*" OR TS="occupation*" OR TS="workplace" OR TS="worker*" OR TS="oral" OR TS="orally" OR TS="ingest*" OR TS="gavage" OR TS="diet" OR TS="diets" OR TS="dietary" OR TS="drinking" OR TS="gastr*" OR TS="intestin*" OR TS="gut" OR TS="sensitiz*" OR TS="abort*" OR TS="abnormalit*" OR TS="embryo*" OR TS="cleft*" OR TS="fetus*" OR TS="foetus*" OR TS="fetal*" OR TS="foetal*" OR TS="fertil*" OR TS="malform*" OR TS="ovum" OR TS="ova" OR TS="ovary" OR TS="placenta*" OR TS="pregnan*" OR TS="dermal*" OR TS="dermis" OR TS="skin" OR TS="epiderm*" OR TS="cutaneous" OR TS="carcinog*" OR TS="cocarcinog*" OR TS="cancer" OR TS="precancer" OR TS="neoplas*" OR TS="tumor*" OR TS="tumour*" OR TS="oncogen*" OR TS="lymphoma*" OR TS="carcinom*" OR TS="genetox*" OR TS="genotox*" OR TS="mutagen*" OR TS="nephrotox*" OR TS="hepatotox*" OR TS="endocrin*" OR TS="estrogen*" OR TS="androgen*" OR TS="hormon*" OR TS="blood" OR TS="serum" OR TS="urine" OR TS="bone" OR TS="bones" OR TS="skelet*" OR TS="rat" OR TS="rats" OR TS="mouse" OR TS="mice" OR TS="guinea" OR TS="muridae" OR TS="rabbit*" OR TS="lagomorph*" OR TS="hamster*" OR TS="ferret*" OR TS="gerbil*" OR TS="rodent*" OR TS="dog" OR TS="dogs" OR TS="beagle*" OR TS="canine" OR TS="cats" OR TS="feline" OR TS="pig" OR TS="pigs" OR TS="swine" OR TS="porcine" OR TS="monkey*" OR TS="macaque"* OR TS="baboon*" OR TS="marmoset*" OR TS="toxic*" OR TS="adverse" OR TS="poisoning" OR TS="prenatal" OR TS="perinatal" OR TS="postnatal" OR TS="reproduc*" OR TS="steril*" OR TS="teratogen*" OR TS="sperm*" OR TS="neonat*" OR TS="newborn*" OR TS="development*" OR TS="zygote*" OR TS="child" OR TS="children" OR TS="adolescen*" OR TS="infant*" OR TS="wean*" OR TS="offspring" OR TS="age") AND PY=(2016-2017) ((TS="DEP" AND TS="phthalat*") OR TS="1,2-benzenedicarboxylic acid, diethyl ester" OR TS="diethyl 1,2-benzenedicarboxylate" OR TS="diethyl o-phthalate" OR TS="diethyl o-phthalates" OR TS="diethyl phthalate" OR TS="diethyl phthalates" OR TS="di-n-ethyl phthalate" OR TS="di-n-ethyl phthalates" OR TS="ethyl phthalate" OR TS="ethyl phthalates" OR TS="phthalic acid, diethyl ester" OR TS="unimoll da" OR TS="solvanol" OR TS="placidol e" OR TS="phthalol" OR TS="palatinol a" OR TS="neantine" OR TS="anozol") AND (TS="genomics" OR TS="proteomics" OR TS="metabolic profile" OR TS="metabolome" OR TS="metabolomics" OR TS="microarray" OR TS="nanoarray" OR TS="gene expression" OR TS="transcript expression" OR TS="transcriptomes" OR TS="transcriptome" OR TS="phenotype" OR TS="transcription" OR TS="trans-act*" OR TS="transact*" OR TS="trans act*" OR TS="genetic" OR TS="genetics" OR TS="genotype" OR TS="genetic transcription" OR TS="gene transcription" OR TS="gene activation" OR TS="genetic induction" OR TS="reverse transcription" OR

Database (date searched)	Terms
	TS="transcriptional activation" OR TS="transcription factors" OR (TS="biosynthesis" AND (TS="RNA" OR TS="DNA")) OR TS="mRNA" OR TS="messenger RNA" OR TS="transfer RNA" OR TS="peptide biosynthesis" OR TS="protein biosynthesis" OR TS="protein synthesis" OR TS="RT-PCR" OR TS="RTPCR" OR TS="reverse transcriptase polymerase chain reaction" OR TS="DNA sequence") AND PY=(2016-2017)
Web of Science 06/2016 01/2016 06/2015 02/2015 07/2014 8/31/13 10/2012	((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS=chronic OR TS=immun* OR TS=lymph* OR TS=neurotox* OR TS=toxicokin* OR TS=pharmacokin* OR TS=biomarker* OR TS=neurolog* OR TS=subchronic OR TS=pbpk OR TS=epidemiolog* OR TS=acute OR TS=subacute OR TS=ld50)
	((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS=lc50 OR TS=inhal* OR TS=pulmon* OR TS=nasal OR TS=lung* OR TS=respir* OR TS=occupation* OR TS=workplace OR TS=worker* OR TS=oral OR TS=orally OR TS=ingest* OR TS=gavage OR TS=diet OR TS=diets OR TS=dietary OR TS=drinking OR TS=gastr* OR TS=intestin*)
	((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS=gut OR TS=sensitiz* OR TS=abort* OR TS=abnormalit* OR TS=embryo* OR TS=cleft* OR TS=fetus* OR TS=foetus* OR TS=fetal* OR TS=foetal* OR TS=fertil* OR TS=malform* OR TS=ovum OR TS=ova OR TS=ovary OR TS=placenta* OR TS=pregnan*)
	((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS=dermal* OR TS=dermis OR TS=skin OR TS=epiderm* OR TS=cutaneous OR TS=carcinog* OR TS=cocarcinog* OR TS=cancer OR TS=precancer OR TS=neoplas* OR TS=tumor* OR TS=tumour* OR TS=oncogen* OR TS=lymphoma* OR TS=carcinom* OR TS=genetox* OR TS=genotox* OR TS=mutagen* OR TS=nephrotox* OR TS=hepatotox* OR TS=endocrin* OR TS=estrogen* OR TS=androgen*)
	((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS=hormon* OR TS=blood OR TS=serum OR TS=urine OR TS=bone OR TS=bones OR TS=skelet* OR TS=rat OR TS=rats OR TS=mouse)

Database (date searched)	Terms
	((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS=mice OR TS=guinea OR TS=muridae OR TS=rabbit* OR TS=lagomorph* OR TS=hamster* OR TS=ferret* OR TS=gerbil* OR TS=rodent* OR TS=dog OR TS=dogs OR TS=beagle* OR TS=canine OR TS=cats OR TS=feline OR TS=pig OR TS=pigs OR TS=swine OR TS=porcine OR TS=monkey* OR TS=macaque* OR TS=baboon* OR TS=marmoset* OR TS=toxic* OR TS=adverse OR TS=poisoning)
	((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS=prenatal OR TS=perinatal OR TS=postnatal OR TS=reproduc* OR TS=steril* OR TS=teratogen* OR TS=sperm* OR TS=neonat* OR TS=newborn* OR TS=development* OR TS=zygote* OR TS=child OR TS=children OR TS=adolescen* OR TS=infant* OR TS=wean* OR TS=offspring OR TS=age)
	-omics search
	2 ((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS="Genomics" OR TS="Proteomics" OR TS="Metabolic Profile" OR TS="Metabolome" OR TS="Metabolomics" OR TS="Microarray" OR TS="Nanoarray")
	11 ((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS="Gene expression" OR TS="Transcript expression" OR TS="transcriptomes" OR TS="transcriptome" OR TS="Phenotype" OR TS="Transcription" OR TS="Trans-act*" OR TS="transact*" OR TS="trans act*" OR TS=genetic OR TS="genetics" OR TS="genotype")
	4 ((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS="Genetic transcription" OR TS="Gene transcription" OR TS="Gene Activation" OR TS="Genetic induction" OR TS="Reverse transcription" OR TS="Transcriptional activation" OR TS="Transcription factors" OR (TS="Biosynthesis" AND (TS=RNA OR TS=DNA)) OR TS="mRNA")
	6 ((TS=DEP AND TS=phthalat*) OR (TS="1,2-Benzenedicarboxylic acid, diethyl ester" OR TS="Diethyl 1,2-benzenedicarboxylate" OR TS="Diethyl o-phthalate" OR TS="Diethyl phthalate" OR TS="Di-n-ethyl phthalate" OR TS="Ethyl phthalate" OR TS="Phthalic acid, diethyl ester")) AND (TS="messenger RNA" OR TS="transfer RNA" OR TS="peptide biosynthesis" OR TS="protein biosynthesis" OR TS="protein synthesis" OR TS="RT-PCR" OR TS="RTPCR" OR TS="Reverse Transcriptase Polymerase Chain Reaction" OR TS="DNA sequence")

Database (date searched)	**Terms**
ToxLine 02/2017 07/2017	@SYN1+@AND+@OR+("diethyl+phthalate"+"diethyl+phthalates"+"ethyl+phthalate"+"ethyl+phthalates"+"1,2-benzenedicarboxylic+acid,+diethyl+ester"+"diethyl+1,2-benzenedicarboxylate"+"diethyl+o-phthalate"+"diethyl+o-phthalates"+"di-n-ethyl+phthalate"+"di-n-ethyl+phthalates"+"phthalic+acid,+diethyl+ester"+"unimoll+da"+solvanol+"placidol+e"+phthalol+"palatinol+a"+neantine+anozol+@TERM+@rn+84-66-2)+@RANGE+yr+2012+2017+@NOT+@org+pubmed+@NOT+@org+pubdart+@NOT+@org+"nih+reporter"
ToxLine 11/2012	@OR+("diethyl phthalate"+"unimoll da"+solvanol+"placidol e"+phthalol+"palatinol a"+neantine+"ethyl phthalate"+anozol+@term+@rn+84-66-2)+@not+@org+pubmed+pubdart+crisp
	@term+@rn+84-66-2+@AND+@org+tscats
TSCATS2, TSCA recent notices 08/2013 10/2012	84-66-2
	84-66-2 (8E OR FYI) TSCA
Toxcenter 03/2012 NOTE: took all non caplus items and caplus with synonyms in titles only, sequence"Dupicates were removed; results were date limited to avoid extensive overlap with Toxline	((84-66-2) not (patent/dt OR tscats/fs)) and (chronic OR immunotox? OR neurotox? OR toxicokin? OR biomarker? OR neurolog? OR pharmacokin? OR subchronic OR pbpk OR epidemiology/st,ct, it) OR acute OR subacute OR ld50# OR lc50# OR (toxicity OR adverse OR poisoning)/st,ct,it OR inhal? OR pulmon? OR nasal? OR lung? OR respir? OR occupation? OR workplace? OR worker? OR oral OR orally OR ingest? OR gavage? OR diet OR diets OR dietary OR drinking(w)water OR (maximum and concentration? and (allowable OR permissible)) OR (abort? OR abnormalit? OR embryo? OR cleft? OR fetus? OR foetus? OR fetal? OR foetal? OR fertil? OR malform? OR ovum OR ova OR ovary OR placenta? OR pregnan? OR prenatal OR perinatal? OR postnatal? OR reproduc? OR steril? OR teratogen? OR sperm OR spermac? OR spermag? OR spermati? OR spermas? OR spermatob? OR spermatoc? OR spermatog? OR spermatoi? OR spermatol? OR spermator? OR spermatox? OR spermatoz? OR spermatu? OR spermi? OR spermo? OR neonat? OR newborn OR development OR developmental? OR zygote? OR child OR children OR adolescen? OR infant OR wean? OR offspring OR age(w)factor? OR dermal? OR dermis OR skin OR epiderm? OR cutaneous? OR carcinog? OR cocarcinog? OR cancer? OR precancer? OR neoplas? OR tumor? OR tumour? OR oncogen? OR lymphoma? OR carcinom? OR genetox? OR genotox? OR mutagen? OR genetic(w)toxic? OR nephrotox? OR hepatotox? OR endocrin? OR estrogen? OR androgen? OR hormon?) AND ("1,2-Benzenedicarboxylic acid, 1,2-diethyl ester"/ti OR "1,2-Benzenedicarboxylic acid, diethyl ester"/ti OR Anozol/ti OR "Diethyl 1,2-benzenedicarboxylate"/ti OR "Diethyl o-phenylenediacetate"/ti OR "Diethyl o-phthalate"/ti OR "Diethyl phthalate"/ti OR "Di-n-ethyl phthalate"/ti OR "DPX-F5384"/ti OR "Estol 1550"/ti OR "Ethyl phthalate"/ti OR Neantine/ti OR "o-Benzenedicarboxylic acid diethyl ester"/ti OR "o-Bis(ethoxycarbonyl)benzene"/ti OR "Palatinol A"/ti OR "Phthalic acid, diethyl ester"/ti OR Phthalol/ti OR "Placidol E"/ti OR Solvanol/ti OR (DEP/ti AND (phthalate/ti OR phthalates/ti))
	–omics search

Database (date searched)	Terms
	("Computational biology" OR "Bio-Informatics" OR Bioinformatics OR ("Molecular Biology" AND Computational) OR Informatics OR ("Information Science" AND Medical))
	Genomics OR Proteomics OR "Metabolic Profile" OR "Metabolome" OR "Metabolomics" OR "Microarray" OR "Nanoarray"
	"Gene expression" OR "Transcript expression" OR transcriptomes OR transcriptome OR Phenotype OR Transcription OR Trans-act? OR
	transact? OR trans()act? OR genetic OR genetics OR genotype
	"Systems biology" OR ("Biological systems" AND (monit? OR data OR analysis))
	(Genetic transcription OR "Gene transcription" OR "Gene Activation" OR "Genetic induction" OR "Reverse transcription" OR "Transcriptional activation" OR "Transcription factors" OR (Biosynthesis AND (RNA OR DNA)))
	mRNA OR "messenger RNA" OR "transfer RNA" OR "peptide biosynthesis" OR "protein biosynthesis" OR "protein synthesis" OR RT-PCR OR RTPCR OR "Reverse Transcriptase Polymerase Chain Reaction" OR "DNA

Appendix Table [SEQ Appendix_Table * ARABIC]. Processes used to augment the search of core databases for DEP

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
Manual search of citations from regulatory documents	NICNAS (National Industrial Chemicals Notification and Assessment Scheme). (2008). Existing chemical hazard assessment report. Diethyl phthalate. National Industrial Chemicals Notification and Assessment Scheme.http://www.nicnas.gov.au/Industry/Existing_Chemicals/Phthalate_Hazard_Assessments/DEP%20hazard%20assessment%2030-4-07.pdf.	5/2013	10 citations added
	ATSDR (Agency for Toxic Substances and Disease Registry). (1995). Toxicological profile for diethyl phthalate. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service.	5/2013	4 citations added
	WHO (World Health Organization). (2003). Concise International Chemical Assessment Document 52: Diethyl phthalate. Geneva. http://www.who.int/ipcs/publications/cicad/en/cicad52.pdf.	5/2013	2 citations added
	U.S. Consumer Product Safety Commission (2010). Toxicity review for diethyl phthalate (DEP).	3/2017	11 citations added

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	Federal Institute for Occupational Safety and Health/Directorate General of Health (2014). Substance evaluation report – DEP.	3/2017	13 citations added
	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin/Directorate-General of Health (2015). Substance evaluation conclusion document as required by REACH Article 48 for diethyl phthalate (DEP).	3/2017	0 citations added
Web of Science, forward search	Jones, HB; Garside, DA; Liu, R; et al. (1993) The influence of phthalate esters on Leydig cell structure and function in vitro and in vivo. Exp Mol Pathol 58:179–193.	6/2013 3/2017	4 citations added 8 citations added
	Shiraishi, K; Miyata, K; Houshuyama, S. (2006) Subacute oral toxicity study of diethylphthalate based on the draft protocol for "Enhanced OECD Test Guideline no. 407". Arch Toxicol. 80: 10-16.	6/2013	0 citations added
	Field, EA; Price, CJ; Sleet, RB; et al. (1993) Developmental toxicity evaluation of diethyl and dimethyl phthalate in rats. Teratology, Jul; 48 (1): 33-44.	6/2013 3/2017	2 citations added 0 citations added
	Swan, SH. (2008). Environmental phthalate exposure in relation to reproductive outcomes and other health endpoints in humans. Environmental Research 108(2): 177-184.	6/2013	10 citations added
	Pereira, C; Mapuskar, K; Rao, CV. (2007) Chronic toxicity of diethyl phthalate--A three generation lactational and gestational exposure study on male Wistar rats. Envir Toxicol and Pharma 23:319–327.	6/2013	0 citations added
	Fujii, S; Yabe, K; Furukawa, M; et al. (2005) A two-generation reproductive toxicity study of diethyl phthalate (DEP) in rats. J Toxicol Sci 30:97–116.	3/2017	3 citations added
	Fisher, JS. (2004) Environmental anti-androgens and male reproductive health: focus on phthalates and testicular dysgenesis syndrome. Reproduction 127:305–315.	3/2017	17 citations added
Backward search, Web of Science or manual	Jones, HB; Garside, DA; Liu, R; et al. (1993) The influence of phthalate esters on Leydig cell structure and function in vitro and in vivo. Exp Mol Pathol 58:179–193.	6/2013	1 citations added
	Shiraishi, K; Miyata, K; Houshuyama, S. (2006) Subacute oral toxicity study of diethylphthalate based on the draft protocol for "Enhanced OECD Test Guideline no. 407". Arch Toxicol. 80: 10-16.	6/2013	0 citations added

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	Field, EA; Price, CJ; Sleet, RB; et al. (1993) Developmental toxicity evaluation of diethyl and dimethyl phthalate in rats. Teratology, Jul; 48 (1): 33-44.	6/2013	2 citations added
	Swan SH. (2008). Environmental phthalate exposure in relation to reproductive outcomes and other health endpoints in humans. Environmental Research 108(2): 177-184.	6/2013	6 citations added
	Pereira, C; Mapuskar, K; Rao, CV. (2007) Chronic toxicity of diethyl phthalate--A three generation lactational and gestational exposure study on male Wistar rats. Envir Toxicol and Pharma 23:319–327.	6/2013	4 citations added
	Fujii, S; Yabe, K; Furukawa, M; et al. (2005) A two-generation reproductive toxicity study of diethyl phthalate (DEP) in rats. J Toxicol Sci 30:97–116.	3/2017	2 citations added
	Gao, HT; et al. (2017) Effects of six priority controlled phthalate esters with long-term low-dose integrated exposure on male reproductive toxicity in rats. Food and Chemical Toxicology 101:94-104.	3/2017	3 citations added
	Fisher, JS. (2004) Environmental anti-androgens and male reproductive health: focus on phthalates and testicular dysgenesis syndrome. Reproduction 127:305–315.	3/2017	0 citations added
References obtained during the assessment process	DEP references in previous assessment or previously added to the HERO project page		47 citations added

Background Check	Searched a combination of CASRNs and synonyms on the following databases: ATSDR ([HYPERLINK "http://www.atsdr.cdc.gov/substances/index.asp"]) CalEPA (Office of Environmental Health Hazard Assessment) ([HYPERLINK "http://www.oehha.ca.gov/risk.html"]) eChemPortal ([HYPERLINK "http://www.echemportal.org/echemportal/participant/page.action?pageID=9"]) EPA Acute Exposure Guideline Levels ([HYPERLINK "http://www.epa.gov/oppt/aegl/pubs/chemlist.htm"]) EPA – IRISTrack/New Assessments and Reviews ([HYPERLINK "http://cfpub.epa.gov/ncea/iris/search/"])[a] EPA NSCEP ([HYPERLINK "http://www.epa.gov/ncepihom/"]) EPA RfD/RfC and CRAVE meeting notes[b] EPA Science Inventory ([HYPERLINK "http://cfpub.epa.gov/si/"]) Federal Docket ([HYPERLINK "http://www.regulations.gov"]) Health Canada First Priority List Assessments ([HYPERLINK "http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl1-lsp1/index-eng.php"]) Health Canada Second Priority List Assessments ([HYPERLINK "http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl2-lsp2/index-eng.php"]) IARC ([HYPERLINK "http://monographs.iarc.fr/ENG/Classification/index.php"])[a] IPCS INCHEM ([HYPERLINK "http://www.inchem.org/"]) ITER (TERA database) ([HYPERLINK "http://www.tera.org/iter/"])[a] NAS via NAP ([HYPERLINK "http://www.nap.edu/"]) NCI ([HYPERLINK "http://www.cancer.gov"]) NCTR	10/2012 2/2017	1 citations added 8 citations added

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	([HYPERLINK "http://www.fda.gov/AboutFDA/CentersOffices/OC/OfficeofScientificandMedicalPrograms/NCTR/default.htm"])[c] NIEHS ([HYPERLINK "http://www.niehs.nih.gov/"]) NIOSHTIC 2 ([HYPERLINK "http://www2a.cdc.gov/nioshtic-2/"]) NTP - RoC, status, results, and management reports ([HYPERLINK "https://seek.niehs.nih.gov/texis/search/?mode=&opts=&pr=internet-all&dropXSL=html&sq=&prox=page&rorder=750&rprox=750&rdfreq=0&rwfreq=0&rlead=1000&rdepth=31&sufs=1&order=r&query=&mu=National+Toxicology+Program"])[a] WHO assessments – CICADS, EHC ([HYPERLINK "http://www.who.int/ipcs/assessment/en/"])		
	ACGIH (Hardcopy TLV booklet and documentation) AIHA WEELs ([HYPERLINK "http://www.tera.org/OARS/WEEL.html"]) ERPGs ([HYPERLINK "https://www.aiha.org/get-involved/AIHAGuidelineFoundation/EmergencyResponsePlanningGuidelines/Documents/2014%20ERPG%20Values.pdf"]) CalEPA Drinking Water Notification Levels ([HYPERLINK "http://www.swrcb.ca.gov/drinking_water/certlic/drinkingwater/NotificationLevels.shtml"]) OEHHA Toxicity Criteria Database ([HYPERLINK "http://www.oehha.ca.gov/tcdb/index.asp"]) Biomonitoring California-Priority Chemicals ([HYPERLINK "http://www.biomonitoring.ca.gov/chemicals/designated-chemicals"]) Biomonitoring California-Designated Chemicals ([HYPERLINK "http://www.biomonitoring.ca.gov/chemicals/designated-chemicals"])	2/2017	

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	OEHHA Fact Sheets ([HYPERLINK "http://www.oehha.ca.gov/public_info/facts/index.html"]) Non-cancer health effects Table (RELs) [HYPERLINK "http://www.oehha.ca.gov/air/allrels.html"] and Cancer Potency Factors (Appendix A and AppendixB) [HYPERLINK "http://www.oehha.ca.gov/air/hot_spots/tsd052909.html"] CHRIP ([HYPERLINK "http://www.safe.nite.go.jp/english/db.html"]) CPSC ([HYPERLINK "http://www.cpsc.gov"]) ECETOC publications ([HYPERLINK "http://www.ecetoc.org/publications"]) ECHA General site [HYPERLINK "http://echa.europa.eu/information-on-chemicals"] info on Registered Substances ([HYPERLINK "http://echa.europa.eu/information-on-chemicals/registered-substances"]) Information from the Existing Substances Regulation (ESR) including Final Risk Assessments ([HYPERLINK "http://echa.europa.eu/information-on-chemicals/information-from-existing-substances-regulation"]) Opinions of the Committee for Risk Assessment on proposals for harmonised classification and labelling [HYPERLINK "https://echa.europa.eu/opinions-of-the-committee-for-risk-assessment-on-proposals-for-harmonised-classification-and-labelling"] Opinions of the RAC adopted under specific ECHA's Executive Director requests [HYPERLINK "https://echa.europa.eu/about-us/who-we-are/committee-for-risk-assessment/opinions-of-the-rac-adopted-under-specific-echa-s-executive-director-requests"] PACT – RMOA and hazard assessment activities [HYPERLINK "https://echa.europa.eu/addressing-		

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	chemicals-of-concern/substances-of-potential-concern/pact?p_p_id=viewsubstances_WAR_echarevsubstanceportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_pos=1&p_p_col_count=2&_viewsubstances_WAR_echarevsubstanceportlet_cur=1&_viewsubstances_WAR_echarevsubstanceportlet_delta=50&_viewsubstances_WAR_echarevsubstanceportlet_keywords=&_viewsubstances_WAR_echarevsubstanceportlet_advancedSearch=false&_viewsubstances_WAR_echarevsubstanceportlet_andOperator=true&_viewsubstances_WAR_echarevsubstanceportlet_orderByCol=synonymDynamicField_504&_viewsubstances_WAR_echarevsubstanceportlet_orderByType=asc"] Substance evaluation – CoRAP [HYPERLINK "https://echa.europa.eu/information-on-chemicals/evaluation/community-rolling-action-plan/corap-table"] European Union Risk Assessment Reports [HYPERLINK "https://ec.europa.eu/jrc/en/publications-list"] Environment Canada – Search entire site ([HYPERLINK "http://www.ec.gc.ca/default.asp?lang=En&n=ECD35C36"]) Health Canada – Search entire site [HYPERLINK "http://www.hc-sc.gc.ca/index-eng.php"] Toxic Substances Managed Under CEPA ([HYPERLINK "http://www.ec.gc.ca/toxiques-toxics/Default.asp?lang=En&n=98E80CC6-1"]) Final Assessments ([HYPERLINK "http://www.ec.gc.ca/lcpe-cepa/default.asp?lang=En&xml=09F567A7-B1EE-1FEE-73DB-8AE6C1EB7658"]) Draft Assessments ([HYPERLINK "http://www.ec.gc.ca/lcpe-cepa/default.asp?lang=En&xml=6892C255-5597-C162-95FC-4B905320F8C9"])		

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	Health Canada Drinking Water Documents [HYPERLINK "http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/index-eng.php" \l "tech_doc"] Health Canada Indoor Air Quality Guidelines [HYPERLINK "http://healthycanadians.gc.ca/healthy-living-vie-saine/environment-environnement/air/guidelines-lignes-directrices-eng.php"] EPA CDAT [HYPERLINK "http://java.epa.gov/oppt_chemical_search/"] OPP [HYPERLINK "http://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1"] FDA ([HYPERLINK "http://www.fda.gov/"]) Japan Existing Chemical Data Base (JECDB) [HYPERLINK "http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp"] NIOSH ([HYPERLINK "http://www.cdc.gov/niosh/topics/"]) NTP 14th Report On Carcinogens: ([HYPERLINK "http://ntp.niehs.nih.gov/?objectid=03C9AF75-E1BF-FF40-DBA9EC0928DF8B15"]) OECD HPV/SIDS/IUCLID (cross-check with eChem) [HYPERLINK "http://webnet.oecd.org/hpv/ui/Search.aspx"] OSHA ([HYPERLINK "http://www.osha.gov/dts/chemicalsampling/toc/toc_chemsamp.html"]) RTECS ([HYPERLINK "http://ccinfoweb.ccohs.ca/rtecs/search.html"]) WHO Air quality guidelines ([HYPERLINK "http://www.who.int/phe/health_topics/outdoorair/outdoorair_aqg/en/"]) Indoor air quality guidelines ([HYPERLINK "http://www.euro.who.int/en/health-topics/environment-and-health/air-		

System Used	Selected Key Reference(s) or Sources	Date	Additional References Identified
	quality/publications/2010/who-guidelines-for-indoor-air-quality-selected-pollutants"]) Drinking water quality guidelines ([HYPERLINK "http://www.who.int/water_sanitation_health/publications/2011/dwq_guidelines/en/"])		

[a]Website updated
[b]Not searched 2/2017 since it's not updated·
[c]Searched as part of FDA for 2/2017